Epoch of reionization 21 cm forecasting from MCMC-constrained semi-numerical models

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **468,** 122–139 (2017) doi:10.1093/mnras/stx420
Advance Access publication 2017 February 17

Epoch of reionization 21 cm forecasting from MCMC-constrained semi-numerical models

Sultan Hassan,[1]★ Romeel Davé,[1,2,3] Kristian Finlator[4] and Mario G. Santos[1,5]

[1] *University of the Western Cape, Bellville, Cape Town 7535, South Africa*
[2] *South African Astronomical Observatories, Observatory, Cape Town 7925, South Africa*
[3] *African Institute for Mathematical Sciences, Muizenberg, Cape Town 7945, South Africa*
[4] *New Mexico State University, Las Cruces, NM 88003, USA*
[5] *SKA SA, The Park, Park Road, Pinelands 7405, South Africa*

Accepted 2017 February 15. Received 2017 February 15; in original form 2016 December 17

ABSTRACT

The recent low value of Planck Collaboration XLVII integrated optical depth to Thomson scattering suggests that the reionization occurred fairly suddenly, disfavouring extended reionization scenarios. This will have a significant impact on the 21 cm power spectrum. Using a semi-numerical framework, we improve our model from instantaneous to include time-integrated ionization and recombination effects, and find that this leads to more sudden reionization. It also yields larger H II bubbles that lead to an order of magnitude more 21 cm power on large scales, while suppressing the small-scale ionization power. Local fluctuations in the neutral hydrogen density play the dominant role in boosting the 21 cm power spectrum on large scales, while recombinations are subdominant. We use a Monte Carlo Markov chain approach to constrain our model to observations of the star formation rate functions at $z = 6$, 7, 8 from Bouwens et al., the Planck Collaboration XLVII optical depth measurements and the Becker & Bolton ionizing emissivity data at $z \sim 5$. We then use this constrained model to perform 21 cm forecasting for Low Frequency Array, Hydrogen Epoch of Reionization Array and Square Kilometre Array in order to determine how well such data can characterize the sources driving reionization. We find that the Mock 21 cm power spectrum alone can somewhat constrain the halo mass dependence of ionizing sources, the photon escape fraction and ionizing amplitude, but combining the Mock 21 cm data with other current observations enables us to separately constrain all these parameters. Our framework illustrates how the future 21 cm data can play a key role in understanding the sources and topology of reionization as observations improve.

Key words: cosmology: theory – dark ages, reionization, first stars – early Universe – galaxies: formation – galaxies:evolution – galaxies: high redshift.

1 INTRODUCTION

The redshifted 21 cm neutral hydrogen line from the epoch of reionization (EoR) provides numerous astrophysical and cosmological information about the formation and evolution of the first stars and galaxies (Barkana & Loeb 2001). Many ongoing and forthcoming experiments such as the Low Frequency Array (LOFAR),[1] the Hydrogen Epoch of Reionization Array (HERA)[2] and the Square Kilometre Array (SKA-Low)[3] are devoted to observing the dense neutral hydrogen gas that traces the cosmic web at redshifts beyond 7. While current experiments only yield upper limits to the measurements of the 21 cm power spectrum, these future experiments are likely to provide a detection in the near future. It is thus important to develop robust and comprehensive theoretical models that can utilize such information, along with observations from other wavelengths and facilities, in order to optimally constrain the physical processes driving reionization.

The recent low value of Planck Collaboration XLVII (2016b) integrated optical depth to Thomson scattering suggests that the EoR may have occurred more suddenly, and at much later times, than what was previously believed (Hinshaw et al. 2013). The low value of $\tau = 0.058 \pm 0.012$ prefers EoR models with late onset and shorter duration. This, in turn, is expected to have a significant impact on the expected 21 cm signal and its evolution. Proper modelling of the sources and sinks of ionizing photons during the EoR is required to accurately model the H II bubbles and study their sizes

★ E-mail: sultanier@gmail.com
[1] http://www.lofar.org/
[2] http://reionisation.org
[3] https://www.skatelescope.org

and distributions. Doing so will enable us to connect the observed 21 cm power spectrum with the physical properties of the sources and sinks of ionizing photons during the EoR.

There are several major challenges to modelling the EoR and its redshifted 21 cm signal, driven by the requirements for accurately modelling the power spectrum of H I on large scales. These requirements include (i) large volumes ($\sim$500 Mpc) in order to capture the large-scale H I fluctuations that will be detected in upcoming 21 cm observations; (ii) high resolution that is sufficient to resolve the ionizing sources and self-shielding systems on sub-kpc scales (Iliev et al. 2015); and (iii) accurate tracking of the ionizing radiation and other feedback processes from the sub-kpc up to Mpc scales. For these reasons, self-consistently simulating the EoR represents an immense computational challenge that no current model has been able to fully meet.

None the less, great progress has been made in simulating the EoR on both small and large scales. Hydrodynamic simulations that self-consistently incorporate radiative transfer (Gnedin 2000, 2014; Pawlik & Schaye 2008; Finlator, Özel & Davé 2009; Finlator et al. 2013; Katz et al. 2016) have sufficient resolution to model the sources of reionization direction, and to propagate the emitted radiation through the intergalactic medium (IGM) with minimal physical assumptions. However, owing to computational limitations, they are currently restricted to volumes smaller than $\sim$20 Mpc in order to resolve all atomically cooling haloes. An alternative approach is to post-process simulated density fields with radiative transfer (Razoumov et al. 2002; Mellema et al. 2006; McQuinn et al. 2007; Thomas et al. 2009; Iliev et al. 2014; Bauer et al. 2015), which allows access to larger volumes but does not self-consistently account for thermal, ionization and chemical feedback effects on galaxy formation. Semi-analytical EoR models (Mitra, Choudhury & Ferrara 2011; Mitra, Ferrara & Choudhury 2013) are very successful in studying and constraining the globally averaged astrophysical quantities and parameters during EoR (Mitra, Choudhury & Ferrara 2012, 2015) based on current observations, but lack the dynamic range to study 21 cm fluctuations. Finally, on the very largest scales, semi-numerical models (Mesinger & Furlanetto 2007; Zahn et al. 2007; Choudhury, Haehnelt & Regan 2009; Santos et al. 2010) based on quasi-linear density evolution with coarse modelling of the source population are able to access volumes sufficient to make 21 cm predictions relevant to upcoming observations, but must employ simple parametrized approximations for the source and sink populations. None the less, such semi-numerical models, with appropriate tuning, can reproduce similar reionization histories as obtained by full radiative transfer simulations (Zahn et al. 2011; Majumdar et al. 2014).

Semi-numerical models are most ideally suited for studying the large-scale ($\gtrsim$1 Mpc) 21 cm power spectrum that will be measured with upcoming radio facilities, but they make many simplifying assumptions. In particular, they must assume parametrizations for the relationship between halo mass and ionizing luminosity, and the relationship between the large-scale density field and the recombination rate that emerges from small-scale clumping. Also, current semi-numerical codes treat a single cell as either fully neutral or fully ionized; hence, they must choose some condition to assign that cell as ionized. This third condition is an algorithmic choice, but the first two connect to physics, as they provide an opportunity to constrain astrophysical quantities associated with EoR sources and sinks based on 21 cm and other EoR observations.

In Hassan et al. (2016), we focused on improving the physical parametrizations of the source and sink populations in the semi-numerical model SIMFAST21 by employing parametrized results from high-resolution radiative hydrodynamic simulations. This enabled greater physical realism of parametrized source and sink populations compared to previous approaches that had used a linear relationship between halo mass and luminosity, and did not include recombinations. To do this, we obtained parametrizations for the ionization rate $R_{\rm ion}$ and recombination rate $R_{\rm rec}$ as functions of halo mass, overdensity and redshift, extracted from high-resolution radiative transfer hydrodynamic simulation (Finlator et al. 2015, hereafter 6/256-RT) and larger volume hydrodynamic simulation (Davé et al. 2013, hereafter 32/512). We then implemented these parametrizations into SIMFAST21, and identified ionized regions where the ionization rate exceeded the recombination rate. This more realistic modelling replaces the canonical efficiency parameter ζ approach in previous semi-numerical EoR modelling. In particular, we found that the $R_{\rm ion}$ scales superlinearly with halo mass ($R_{\rm ion} \propto M_{\rm h}^{1.4}$) in contrast to the typically assumed linear relationship between the efficiency parameter ζ and halo mass. We showed that using these new parametrizations ($R_{\rm ion}$ and $R_{\rm rec}$) allows us to simultaneously match various EoR key observables with a relatively low escape fraction, independent of halo mass and redshift. We also found that the $R_{\rm ion}$ boosts the small-scale 21 cm power spectrum while $R_{\rm rec}$ suppresses the 21 cm power on large scales during cosmic reionizations.

Hassan et al. (2016) thus improved upon the first two major uncertainties in semi-numerical models, namely the ionization and recombinations. However, this work still assumed an ionization condition based on the *instantaneous* balance between ionizations and recombinations – in other words, if there were instantaneously more ionizations than recombinations, that volume of space was considered fully ionized. However, this is not physically fully accurate, because the excess ionizing photons in such regions still must ionize the neutral hydrogen atoms in that region. The instantaneous criterion thus does not account for partial ionization of a given cell; thus, it underestimates the total number of photons required. In the limiting case where reionization proceeds quickly, this may not be a bad approximation, but ideally we aim to relax this instantaneous assumption. In essence, it is likely that our ionizations were too efficient, which can affect the topology and duration of the EoR along with our constraints on $f_{\rm esc}$.

In this paper, we improve upon our previous ionization condition by tracking the actual number of neutral hydrogen atoms, ionizing photons and recombinations. This leads to a time-dependent ionization condition that is analogous to the well-known ionization balance equation. With this, it turns out that reionization occurs more suddenly, as preferred by the recent Planck Collaboration XLVII (2016b) constraints, but requires a higher escape fraction. We compare this to our previous instantaneous EoR model (Hassan et al. 2016) in terms of their H II bubble sizes, EoR history and 21 cm power spectra.

Ideally, we would like to use the 21 cm power spectra and other observations to provide constraints on the nature of the source population, in particular its relationship to the halo population. In Hassan et al. (2016), we manually constrained the relationship between ionizing emissivity and halo mass versus observations, since we only had one free parameter, namely the escape fraction of ionizing photons. This was because we had fixed the characteristics of the source population based on our radiative hydrodynamic simulations. Here we would like to relax this assumption, and determine how well we can constrain the source population characteristics directly from observations. To do this, we consider a generalized model with three free parameters: the escape photon fraction $f_{\rm esc}$, the ionizing emissivity amplitude $A_{\rm ion}$ ($R_{\rm ion}$ amplitude) and the ionizing

emissivity–halo mass power-law index $C_{\rm ion}$. We note that $C_{\rm ion}$ can represent the power-law mass dependence of either the amplitude $A_{\rm ion}$ or the escape fraction $f_{\rm esc}$; in our current approach, these two quantities are degenerate. To constrain these parameters, we perform a Bayesian Monte Carlo Markov chain (MCMC) search against current EoR observations. We then forecast how these constraints will be improved by upcoming 21 cm observations from LOFAR, HERA and SKA-Low. By considering all such observations, we determine how well we can constrain the EoR source population as characterized by our three free parameters.

This paper is organized as follows. In Section 2, we introduce our previous instantaneous EoR model and our new time-integrated model. We study and compare these models' impact on various EoR observables including the 21 cm power spectrum in Section 3. In Section 4, we create several EoR models to study their effects on the 21 cm power spectrum. In Section 5, we calibrate the time-integrated model to various EoR observations. We perform the 21 cm forecasting in Section 6 and draw our concluding remarks in Section 7.

Throughout this work, we adopt a Λ cold dark matter cosmology in which $\Omega_{\rm M} = 0.3$, $\Omega_\Lambda = 0.7$, $h \equiv H_0/(100\,{\rm km\,s^{-1}\,Mpc^{-1}}) = 0.7$, a primordial power spectrum index $n = 0.96$, an amplitude of the mass fluctuations scaled to $\sigma_8 = 0.8$ and $\Omega_{\rm b} = 0.045$. We quote all results in comoving units, unless otherwise stated.

2 SIMULATIONS

We use a semi-numerical code SIMFAST21 (Santos et al. 2010), which we briefly review here. SIMFAST21 simulation begins by generating the density field from a Gaussian distribution using a Monte Carlo approach. The generated density field will then be dynamically evolved from linear to non-linear regime by applying the Zel'Dovich (1970) approximation. The dark matter haloes are generated using the well-known excursion-set formalism. In the standard SIMFAST21, the ionized regions are identified using the excursion-set formalism based on a constant efficiency parameter ζ. In the original SIMFAST21 code, the ionization condition compares the amount of collapsed dark matter halo $f_{\rm coll}$ to the efficiency parameter ζ – any region will be flagged as ionized if

$$f_{\rm coll} \geq \zeta^{-1}. \tag{1}$$

The efficiency parameter ζ is a model free parameter that can be tuned to match some observations. This condition generates the ionization field, which may be used along with the density field to obtain the 21 cm brightness temperature. We refer the reader to Santos et al. (2010) for more details on this model and the code algorithm.

We now describe our two extensions to SIMFAST21. The first was presented in Hassan et al. (2016), which we review next, and incorporates the ionization and recombination rate parametrizations taken from hydrodynamic simulations, but utilizes an instantaneous ionization condition. We then describe our further extension here in order to improve the ionization condition by tracking the neutral fraction in a time-integrated manner. We will call this the 'instantaneous ionization' and 'time-integrated ionization' models.

2.1 Instantaneous ionization model

As described in Hassan et al. (2016), here we replace the efficiency parameter ζ with direct parametrizations of the ionization rate $R_{\rm ion}$ and recombination rate $R_{\rm rec}$ as functions of halo mass $M_{\rm h}$, overdensity Δ and redshift z, taken from our 6/256-RT and 32/512 simulations. Our best-fitting non-linear ionization rate $R_{\rm ion}$ parametrization takes the following form:

$$\frac{R_{\rm ion}}{M_{\rm h}} = A_{\rm ion}\,(1+z)^{D_{\rm ion}}\,(M_{\rm h}/B_{\rm ion})^{C_{\rm ion}}\,\exp\left(-(B_{\rm ion}/M_{\rm h})^{3.0}\right), \tag{2}$$

where $A_{\rm ion} = 1.08 \times 10^{40}\,{\rm M_\odot^{-1}\ s^{-1}}$, $B_{\rm ion} = 9.51 \times 10^7\,{\rm M_\odot}$, $C_{\rm ion} = 0.41$ and $D_{\rm ion} = 2.28$. Meanwhile, we parametrize the recombination rate as

$$\frac{R_{\rm rec}}{V} = A_{\rm rec}(1+z)^{D_{\rm rec}}\left[\frac{(\Delta/B_{\rm rec})^{C_{\rm rec}}}{1+(\Delta/B_{\rm rec})^{C_{\rm rec}}}\right]^4, \tag{3}$$

where $A_{\rm rec} = 9.85 \times 10^{-24}\ {\rm cm^{-3}\ s^{-1}}$, $B_{\rm rec} = 1.76$, $C_{\rm rec} = 0.82$, $D_{\rm rec} = 5.07$.

Our ionization condition is taken to be

$$f_{\rm esc}\,R_{\rm ion,V} \geq R_{\rm rec,V}, \tag{4}$$

where

$$R_{\rm ion,V} = \int_{\rm dn}\int_V \frac{{\rm d}n}{{\rm d}M_{\rm h}} R_{\rm ion}(M_{\rm h}, z)\,{\rm d}M_{\rm h}\,{\rm d}V,$$

and

$$R_{\rm rec,V} = \int_V R_{\rm rec}(\Delta, z)\,{\rm d}V.$$

In above expressions, the $f_{\rm esc}$ is the photon escape fraction, V is the spherical region volume specified by the excursion-set formalism and n is the number density of haloes. With these volume integrals, the $R_{\rm ion,\,V}$ represents the total ionization rate from all sources and $R_{\rm rec,\,V}$ is the maximum recombination rate in that volume V. Cells in a given volume V satisfying this criterion (equation 4) are considered fully ionized, otherwise they are fully neutral.

Using this model, it has been shown in Hassan et al. (2016) that one can match simultaneously several EoR key observables, such as Planck Collaboration XIII (2016a) optical depth, Becker & Bolton (2013) ionizing emissivity and Fan, Carilli & Keating (2006) filling factor measurement, by only a constant $f_{\rm esc} = 4$–6 per cent independent of halo mass or redshift. We refer the reader to Hassan et al. (2016) for more details about the model and these new parametrizations.

2.2 Time-integrated ionization model

The model in Hassan et al. (2016, our instantaneous EoR model) has several drawbacks. First, the ionization condition, equation (4), is an instantaneous criterion that compares the escaped ionization rate $f_{\rm esc}R_{\rm ion}$ with the recombination rate $R_{\rm rec}$, instead of comparing the actual numbers of ionizing photons to that of recombinations. Secondly, the ionization condition, equation (4), assumes maximum recombination rate $R_{\rm rec}$ from all cells/regions as if they were fully ionized.

To improve on these, we modify the ionization condition to account for the evolving neutral hydrogen fraction, which allows us to account for the existing number of hydrogen atoms in each region as well as to compute the recombination based on the current ionized fraction. Hence, we now employ a time-dependent integral ionization condition:

$$f_{\rm esc}\,R_{\rm ion,V}\ {\rm d}t \geq \mathbb{R}_{\rm rec,V}\ {\rm d}t + N_{\rm HI,V}, \tag{5}$$

where

$$\mathbb{R}_{\rm rec,V} = \int_{\rm V} x_{\rm H\,II}\,R_{\rm rec}(\Delta, z)\,{\rm d}V,$$

and

$$N_{\rm HI,V} = \int_{\rm V} (1 - x_{\rm H\,II})\, N_{\rm H}\, {\rm d}V.$$

The $x_{\rm H\,II}$ and $N_{\rm H}$ here are the ionization fraction and the total number of hydrogen in cells, respectively. The dt represents the time duration between successive snapshots. We apply this new condition (equation 5) as follows: for each cell, we first compute the integrand of the LHS ($f_{\rm esc} R_{\rm ion} {\rm d}t$) and RHS ($x_{\rm H\,II} R_{\rm rec} {\rm d}t + (1 - x_{\rm H\,II}) N_{\rm H}$). We then apply the excursion-set formalism to perform the spherical volume integral. Once again, the ionization fraction is set to 1 (fully ionized) for cells in volumes that satisfy the ionization condition (equation 5), otherwise they remain fully neutral with zero ionization fraction.

The left-hand side (LHS) of our new ionization condition (equation 5) represents the actual number of escaped ionizing photons being emitted in dt. The first term of the right-hand side (RHS) of this condition is the actual number of recombinations occurring during dt in regions with an ionization fraction of $x_{\rm H\,II}$. This term then tracks the exact number of recombinations even from partially ionized regions with $0 < x_{\rm H\,II} < 1$. This recombination term has no effect at early times of EoR when the Universe is completely neutral, but becomes the dominant sink for ionizing photons at late stages of the EoR (Sobacchi & Mesinger 2014).

The second term of the RHS of equation (5) denotes the total number of neutral hydrogen in the simulation box. At early EoR stages, the escaped ionizing photons (LHS) fight only with the neutral hydrogen term $(1 - x_{\rm H\,II}) N_{\rm H}$. As the EoR proceeds, the neutral hydrogen term becomes less significant, and the recombinations start to play the leading role in forestalling reionization. Hence, this provides a more physically motivated ionization condition, in a similar form to the standard ionization balance equation.

The condition is clearly time dependent, unlike the instantaneous ionization model for which the ionization condition could be evaluated independently at each time-step. Thus, our new scheme can be sensitive to the choice of the time-step dt used to perform the integration. For instance, a larger dt(dz) will result in more ionizing photons and also more recombinations. This then leads to a wrong evolving ionization balance. We have conducted convergence tests to determine that d$z = 0.125$ provides a numerically converged answer (see Appendix A). Our new method thus requires higher computational cost to evolve the ionization state forward in time. We will explore possible variations of the ionization condition from equation (5) in Section 4, to study their impact on the 21 cm power spectrum.

It turns out, as we will show, that the instantaneous ionization condition results in more extended reionization history, while our new time-integrated condition yields more sudden reionization. Next, we will investigate their differences in terms of the EoR history, topology and the 21 cm power spectrum.

3 IMPACT ON EoR OBSERVABLES

We use the Full model from Hassan et al. (2016) as our fiducial 'instantaneous' reionization model. This model uses equation (4) to identify the ionized regions with $f_{\rm esc} = 4$ per cent in a large-volume box of $L = 300$ Mpc and a number of cells $N = 560^3$. This model yields a maximum halo mass of $3.56 \times 10^{12}\,{\rm M}_\odot$ and a minimum halo mass of $1.28 \times 10^{8}\,{\rm M}_\odot$ at $z = 6$. We have shown that this model matches various observations of the EoR including the Planck Collaboration XIII (2016a) optical depth $\tau = 0.066$. Using the same density field boxes and halo catalogues, we run our new time-integrated reionization model with parameters calibrated against various EoR key observations (see Section 5), including the new Planck Collaboration XLVII (2016b) optical depth. The two models are tuned to different τ_e values, but we do our 21 cm comparison at a given neutral fraction since it has been shown that the 21 cm power spectrum shape is more sensitive to the neutral fraction (e.g. see Mesinger & Furlanetto 2007; Zahn et al. 2007). We also verify this later by comparing the instantaneous model power spectra at different redshifts for a fixed neutral fraction in Fig. 5 in Section 3.3. We thus begin by comparing the instantaneous and time-integrated models' differences in terms of their global neutral fraction history.

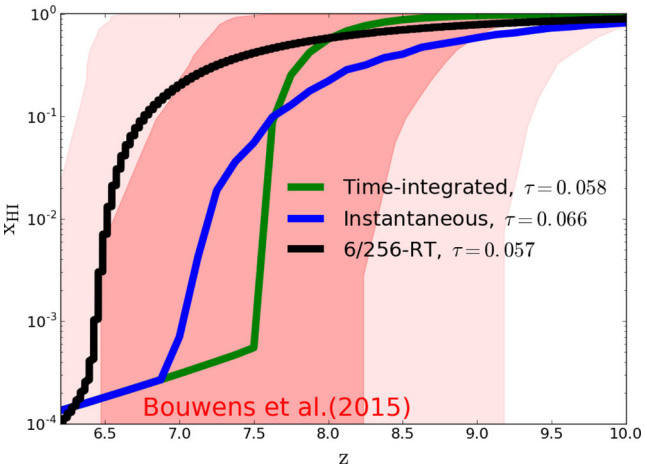

Figure 1. The volume-weighted average neutral fraction evolution as a function of redshift. The time-integrated EoR model ($\tau = 0.058$), the instantaneous EoR model developed in Hassan et al. (2016) ($\tau = 0.066$) and the 6/256-RT simulation ($\tau = 0.057$) are represented by the green, blue and black lines, respectively. The shaded areas are several quasars and Lyα constraints as compiled by Bouwens et al. (2015). It is quite clear that all models are consistent with the observational constraints by Bouwens et al. (2015). Differences between models are explained in the text.

3.1 EoR ionization history

Fig. 1 shows the global reionization history produced by our two fiducial models, instantaneous and time-integrated, compared to the neutral fraction constraints obtained by Bouwens et al. (2015) via a compilation from various observables. We immediately see that the green line showing the new time-integrated ionization condition shows a more sudden transition from fully ionized to fully neutral. Meanwhile, the blue line from our old instantaneous condition results in a more extended reionization epoch. None the less, in both cases, reionization occurs in our two models within observational constraints (light red shaded areas). It is perhaps worth noting that, unlike a few years ago when the canonical redshift for the end of reionization was regarded as $z \sim 6$, current constraints from both observations and models favour the end of reionization to occur at $z \sim 7$ or perhaps a bit higher.

This plot already shows that accounting for the neutral gas through comparing the number of neutral atoms and ionizing photons (equation 5) versus comparing instantaneous rates (equation 4) has a significant impact on the reionization history. The time-integrated model is qualitatively more compatible with the picture that has been suggested by recent Planck Collaboration XLVII (2016b) constraints that favour sudden EoR scenarios. We emphasize the fact that if we tune the time-integrated model optical depth

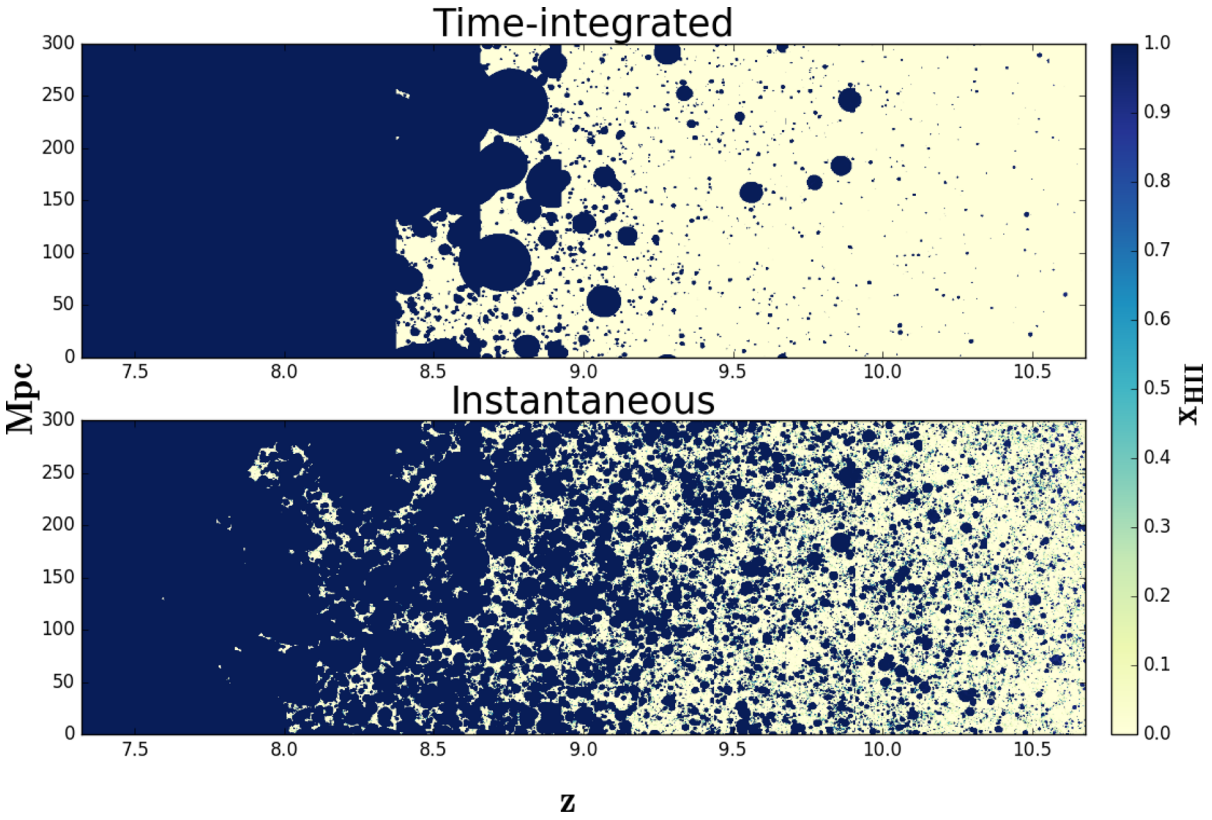

Figure 2. Evolving maps of the neutral fraction from the instantaneous and time-integrated models. Time-integrated EoR model produces large H II bubbles and reionizes the Universe very rapidly, indicating a sudden EoR scenario. Instantaneous EoR model yields small H II bubbles and reionizes the Universe very late, leading to an extended EoR scenario.

to match the instantaneous model optical depth ($\tau = 0.066$; Planck Collaboration XIII 2016a), the time-integrated model will require higher f_{esc} and shift reionization towards higher redshifts, but nevertheless the reionization history shape will remain sudden as shown later in Fig. 9. We will come back to this point later in Section 5.2.2. However, when using the same parameters, the reionization in the time-integrated model is delayed by $\Delta z \sim 0.8$ as compared with that in the instantaneous model.

From Fig. 1, we also see that both models instantaneous and time-integrated reionize the Universe earlier than the 6/256-RT simulation. As discussed before in Hassan et al. (2016), the small box size ($6\ h^{-1}$ Mpc) of 6/256-RT does not capture the large-scale fluctuations that give rise to the most massive haloes that provide a significant fraction of reionizing photons. Hence, at a fixed optical depth, it is expected that the 6/256-RT might reionize the Universe much later than the time-integrated model due to the box size limitations.

An informative way to examine these models is by viewing lightcones, as shown in Fig. 2. These have been constructed by projecting the ionization state within the simulation volume along a specific line of sight, evolving with redshift. Fig. 2 confirms that our previous model (the Full model of Hassan et al. 2016) produces a more extended EoR scenario that corresponds to an early onset and a very late end with a duration of $\Delta z \sim 10$. Unlike the instantaneous model, Fig. 2 also shows that our new model (equation 5) yields a sudden reionization scenario where the EoR starts very late (once $x_{H\,I} < 1.0$) and ends (when $x_{H\,I}$ drops below 10^{-3}) very quickly within a duration of $\Delta z \sim 4$. More strikingly, it also shows that the time-integrated EoR model produces larger ionized bubbles while the instantaneous model yields many ionized bubbles of smaller sizes. This can further be quantified by studying their differences in terms of the ionization field power spectra.

3.2 EoR topology

It is useful to compare the models at a specific neutral fraction, since this best illustrates the difference in topology. Fig. 3 compares our instantaneous and time-integrated models in terms of their ionization maps when the EoR is half-way through, i.e. with a globally averaged $x_{H\,I} \sim 0.5$. These ionization maps show the spatial distribution of the large and small ionized bubbles (black regions) over 300 Mpc scales. However, the excursion-set formalism with its binary structure (fully ionized or fully neutral), along with the large cell size of ∼0.5 Mpc, prevents these models' maps to display the self-shielded regions in the ionized media as seen in Figs 3 and 2. The presence of these self-shielded regions does not affect the 21 cm power spectrum but rather lower the ionization fraction at intermediate densities ($\Delta = 5$–10) as previously shown (see fig. 10 in Hassan et al. 2016). We have also shown in Hassan et al. (2016) that including the sub-clumping effect on scales below our cell size (∼0.5 Mpc) has a minimal effect on the expected signal [see comparison between Full and NoSubClump models in Hassan et al. (2016) for more details].

From Fig. 3, we see that the instantaneous model produces many small H II bubbles more uniformly distributed across the ionization map. This shows that the EoR in the instantaneous model proceeds from small scales, and the ionizing photons are able to reionize locally everywhere. This is because the instantaneous ionization

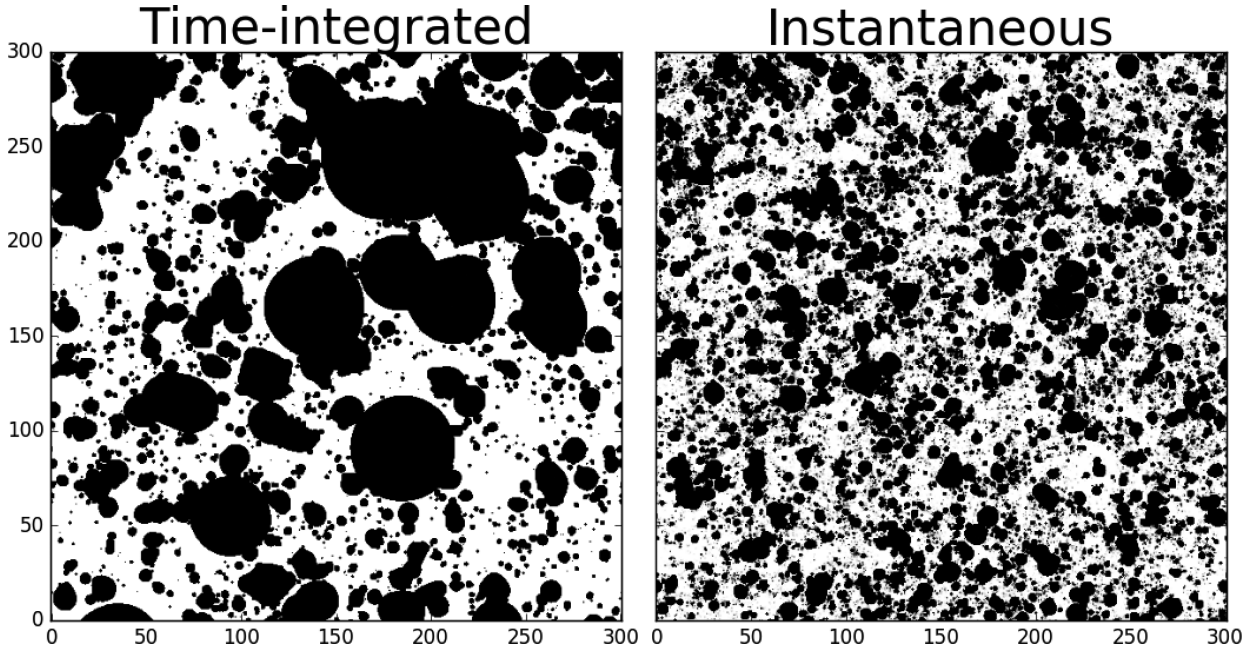

Figure 3. Slice of the ionization box of a size $300 \times 300 \times 0.535\,\mathrm{Mpc}^3$ from the instantaneous and time-integrated models and at $x_{\mathrm{H\,I}} \sim 0.5$. White and black represent neutral and ionized regions, respectively. The self-shielded regions in ionized media are absent due to the binary structure of the excursion-set formalism (assigning 1 to ionized and 0 to neutral cells in regions that satisfy the ionized condition) along with the large cell size of ∼0.5 Mpc.

rate can easily exceed the recombination rate (see equation 4) on small scales when neglecting the local neutral hydrogen content.

In contrast, the time-integrated model ionization map shows very large H II bubbles. This may be explained by interpreting the time-integrated model ionization condition (equation 5). As noted earlier, at high redshifts when the Universe is neutral ($x_{\mathrm{H\,II}} \sim 0$), the recombination term can be neglected. In this case, the time-integrated model only compares the escaped ionizing photons with the total number of neutral hydrogen atoms. This condition dominates until the region becomes partially ionized. At that point, recombinations will start to occur because of the non-zero ionization fraction $x_{\mathrm{H\,II}}$, but still this region is now less neutral, which will allow more rapid ionization. However, different forms of ionization condition yield very different H I fluctuations (see Section 4). In general, the sources and sinks are occurring within the regions that are densest and thus contain the most number of neutral hydrogen atoms. The high density causes the ionizing photons to be ineffective at ionizing local regions, until such time as significant ionizations happen, which then rapidly ionize the surrounding regions.

Fig. 4 shows the ionization field power spectra of our fiducial models at different stages of reionization when the Universe is 25 per cent, 50 per cent and 75 per cent reionized. These neutral fractions correspond to $z = 8.0$, 8.75, 9.5 and $z = 7.75$, 8.0, 8.25 as obtained by the instantaneous and time-integrated models, respectively. This is computed as follows: $\Delta^2_{xx} \equiv k^3/|(2\pi^2\,V)\langle\,|x_{\mathrm{H\,II}}|^2\rangle/x^2_{\mathrm{H\,I}}$ (Hassan et al. 2016).

The time-integrated model produces more power on large scales by 1 to 1.2 orders of magnitude and less power on small scales by a factor of 2–3, at fixed ionization fraction, as compared to the instantaneous model. This is consistent with the qualitative impression from the ionization maps in Fig. 3. We further see that the large-scale ionization power spectrum, obtained by the time-integrated EoR model, peaks at ∼75 Mpc that corresponds to the characteristic size of the H II bubbles as seen in the H II maps in Fig. 3.

The difference, particularly on large scales, is substantial, which shows the importance of accounting for the existing neutral hydrogen content in the ionization condition [i.e. the second term of equation (5)]. Since the fluctuations in the ionization field drive the 21 cm brightness temperature, we expect to see similar differences in the 21 cm power spectra, which we examine next.

3.3 The 21 cm power spectrum

Using the ionization fields of these models, we now compute our EoR key observable that is the 21 cm power spectrum. Assuming that the spin temperature is much higher than the cosmic microwave background (CMB) temperature, the 21 cm brightness temperature takes the following form:

$$\delta T_{\mathrm{b}}(\nu) = 23 x_{\mathrm{H\,I}} \Delta \left(\frac{\Omega_{\mathrm{b}} h^2}{0.02}\right) \sqrt{\frac{1+z}{10}\frac{0.15}{\Omega_{\mathrm{m}} h^2}} \left(\frac{H}{H + \mathrm{d}v/\mathrm{d}r}\right)\ \mathrm{mK}, \tag{6}$$

where $\mathrm{d}v/\mathrm{d}r$ is the comoving gradient of the line-of-sight component of the comoving velocity. Using this equation, it is straightforward to create the 21 cm brightness temperature boxes from which we compute the 21 cm power spectrum as follows: $\Delta^2_{21} \equiv k^3/(2\pi^2\,V)\langle|\delta T_{\mathrm{b}}(k)|^2_k\rangle$.

We first verify that the 21 cm power spectrum is primarily sensitive to the global ionization fraction, while the density field evolution is secondary. Note that the 21 cm fluctuations trace those of the density field only at early times when the Universe is almost neutral. Here we quote results for the instantaneous model, but we expect that this is also valid for other models such as our time-integrated model. We tune the instantaneous model to Planck Collaboration XLVII (2016b) optical depth ($\tau = 0.058$) that yields $x_{\mathrm{H\,I}} = 0.75$, 0.54, 0.277 at $z = 8.75$, 8.00, 7.25, respectively. We then re-tune the model to the Planck Collaboration XIII (2016a) optical depth ($\tau = 0.066$) [similar to our previous Full model in Hassan

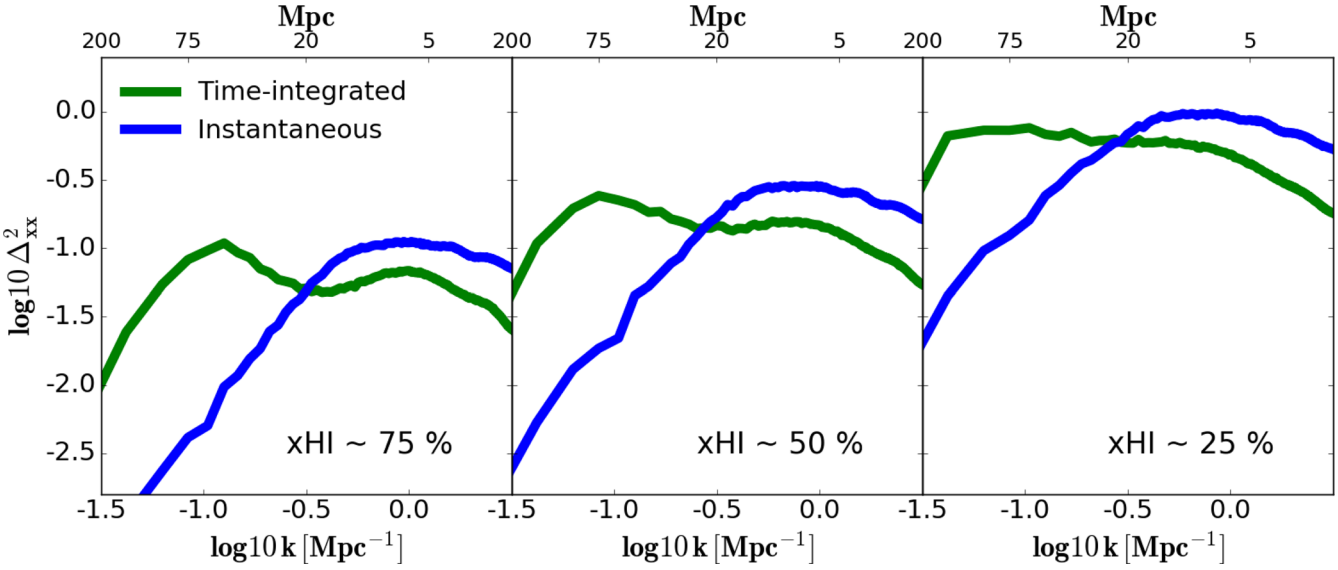

Figure 4. Ionization field power spectrum comparison between the instantaneous (blue) and time-integrated (green) models at different stages of reionization ($x_{\rm H\,I} \sim$ 25 per cent, 50 per cent, 75 per cent).

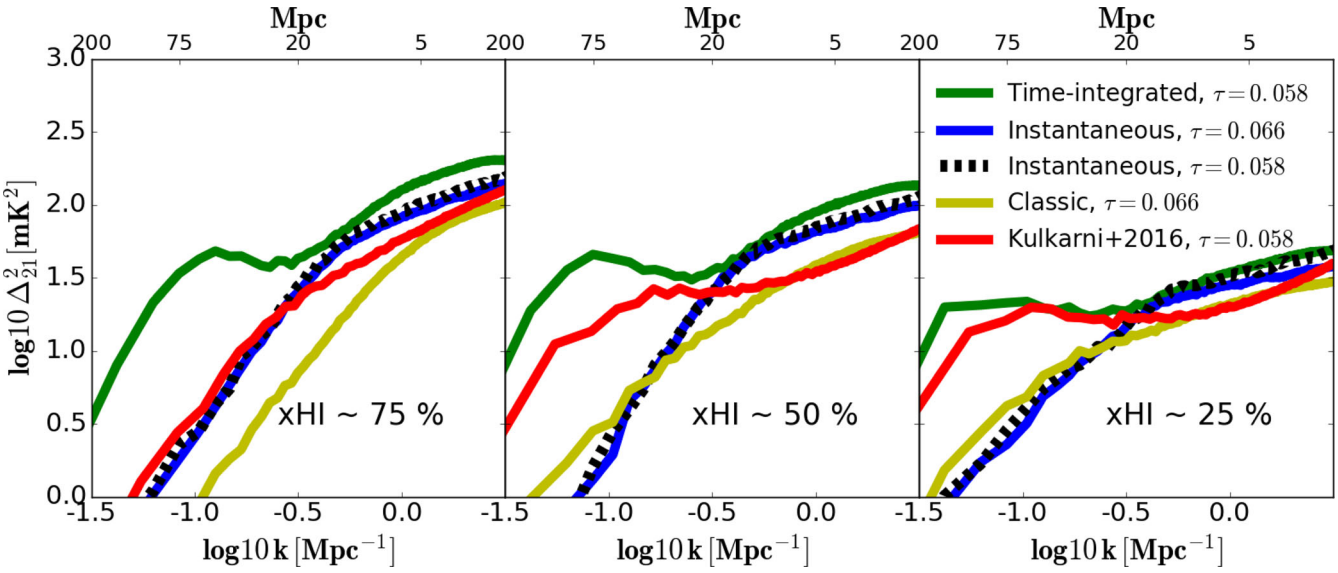

Figure 5. The 21 cm power spectrum comparison between the instantaneous (blue) and time-integrated (green) models at different stages of reionization ($x_{\rm H\,I} \sim$ 25 per cent, 50 per cent, 75 per cent). We compare our 21 cm power spectra with the Very Late model (red) by similar semi-numerical method (Kulkarni et al. 2016) that is calibrated to match Lyα and CMB data. Although our time-integrated and Kulkarni et al. (2016) models correspond to different redshifts, but nevertheless the shape of the 21 cm spectrum is similar, particularly at the intermediate and late stages of reionizations. We also show our classic EoR model (yellow) from Hassan et al. (2016) that adopts the standard efficiency parameter approach similar to Kulkarni et al. (2016) model. Our classic and Kulkarni et al. (2016) models produce similar power on small scales. The instantaneous model, whether it is tuned to Planck Collaboration XLVII (2016b) τ (black dashed) or Planck Collaboration XIII (2016a) τ (solid blue), always produces the exact 21 cm power spectrum at fixed neutral fractions, regardless of the density field contribution from different redshifts. Models are tuned to the two recent *Planck* optical depth values as quoted in the legend.

et al. (2016)] to obtain $x_{\rm H\,I} = 0.73, 0.53, 0.25$ at $z = 9.5, 8.75, 8.0$, respectively. We now compare their difference in the 21 cm power spectrum at these different redshifts for a fixed neutral fraction in Fig. 5. Comparing the solid blue line with black dashed line, we find that the the instantaneous model produces the exact 21 cm power spectrum at a fixed neutral fraction, irrespective of the density field evolution at different redshifts. Hence, we will compare different models at similar neutral fractions, not similar redshifts.

Fig. 5 shows the 21 cm power spectrum of the instantaneous and time-integrated models at neutral fractions of 25 per cent, 50 per cent and 75 per cent. Mimicking the ionization field power spectrum, the time-integrated model produces more power on large scales by 1 to 1.2 orders of magnitude at fixed ionization fraction, as to that of the instantaneous model. Likewise, the time-integrated model also produces slightly more power on small scales by a factor of 1.2–1.5 as compared with the instantaneous EoR model. This difference is less than when comparing the ionization field power spectra, which comes from the contribution of the density field to the 21 cm power spectrum – small regions with high local density (high recombinations) remain neutral, and hence they do not contribute much to the small-scale fluctuations in 21 cm power.

Table 1. Summary of models considered in Section 4 for the 21 cm power spectrum comparison in Fig. 6 from different physical assumptions and treatment to the integrands of RHS integrals in the time-integrated ionization condition, equation (5).

Model class	Neutral hydrogen term	Recombination term
Full-NH-Full-Rrec (FNH-FRrec)	$N_{\rm H}$	$R_{\rm rec}$
Full-NH-Partial-Rrec (FNH-PRrec)	$N_{\rm H}$	$x_{\rm H\,II}\,R_{\rm rec}$
Full-NH-No-Rrec (FNH-NRrec)	$N_{\rm H}$	0
Partial-NH-Full-Rrec (PNH-FRrec)	$(1-x_{\rm H\,II})\,N_{\rm H}$	$R_{\rm rec}$
Partial-NH-Partial-Rrec (PNH-PRrec)	$(1-x_{\rm H\,II})\,N_{\rm H}$	$x_{\rm H\,II}\,R_{\rm rec}$
Partial-NH-No-Rrec (PNH-NRrec)	$(1-x_{\rm H\,II})\,N_{\rm H}$	0
No-NH-Full-Rrec (NNH-FRrec)	0	$R_{\rm rec}$
No-NH-Partial-Rrec (NNH-PRrec)	0	$x_{\rm H\,II}\,R_{\rm rec}$

We also compare our 21 cm power spectra to a similar semi-numerical model by Kulkarni et al. (2016) that has been calibrated with Lyα and CMB data. The semi-numerical models by Kulkarni et al. (2016) adopt the standard efficiency parameter (ζ) approach similar to our classic EoR model (yellow in Fig. 5) from Hassan et al. (2016). We choose to compare with the Very Late model in Kulkarni et al. (2016, red in Fig. 5) that is tuned to match the Planck Collaboration XLVII (2016b) optical depth, consistent with the optical depth produced by our time-integrated model. The ionization histories of Kulkarni et al. (2016) and our time-integrated model are very different even though they obtain ∼50 per cent neutral fraction at the same redshift. For instance, our time-integrated model produces $x_{\rm H\,I} = 0.77, 0.57, 0.25$ at $z = 8.25, 8.0, 7.75$ whereas Kulkarni et al. (2016) model finds $x_{\rm H\,I} = 0.84, 0.59, 0.42$ at $z = 10.0, 8.0, 7.0$. Regardless of this difference in these models' reionization histories, their 21 cm power spectra are generally similar. We find that both models produce a similar shape of the 21 cm power spectrum particularly during the intermediate and final stages of reionization. The minor difference in their amplitudes is due to using our $R_{\rm ion}$–$R_{\rm rec}$ versus the standard ζ approach. This can be clearly seen when comparing our classic EoR model with Kulkarni et al. (2016) model in Fig. 5. We see both models produce the same power on small scales while their difference on large scales might be from the difference in the density field and neutral fractions. This confirms our previous findings that using the non-linear ionization power, via our $R_{\rm ion}$–$R_{\rm rec}$ approach, boosts the 21 cm power spectrum as compared to models adopting the standard efficiency parameter method [classic and Kulkarni et al. (2016) models].

This shows that our time-integrated model, that is calibrated to match various EoR key observables, produces similar 21 cm power spectrum as obtained by other semi-numerical models that have been calibrated to match Lyα and CMB data. The future 21 cm observations might be able to discriminate between these models' power spectra.

In summary, we have compared the instantaneous and time-integrated models in terms of their EoR history, topology and their 21 cm power spectra. We have found that the time-integrated model produces large H I bubbles while the instantaneous model produces more small H I bubbles. The time-integrated model yields a large-scale 21 cm/ionization power spectrum that is higher by one order of magnitude as compared with the instantaneous model. We have seen that the ionization condition (equation 5) results in large H II bubbles that boost the amount of power on large scales. The comparison presented here aims to summarize the differences found between our new (time-integrated) and previous (instantaneous) models. However, previous works by Zahn et al. (2011) and Majumdar et al. (2014) have shown that semi-numerical simulations agree with radiative transfer simulations in terms of their ionization fields and 21 cm power spectra. We leave for future work whether this new model matches radiative transfer simulations.

4 MODEL ASSUMPTION EFFECTS ON THE 21 CM POWER SPECTRUM

The large differences in the 21 cm power spectrum (Fig. 5) between the instantaneous and time-integrated models show that the 21 cm power spectrum is highly sensitive to the physical assumptions used. There are two main differences between these models. First, the ionization condition now accounts for the number of hydrogen atoms, and secondly, the recombination is now done accounting for partial ionization. We believe that our new model is more physically motivated and realistic, but we would like to understand exactly how these changes individually impact the 21 cm power spectra.

We therefore consider the various possible combinations between $R_{\rm rec}$ and $N_{\rm H}$ to create several models with different ionization condition. We also consider models neglecting recombination altogether to analyse the impact of recombinations as we did in Hassan et al. (2016), only with our new time-integrated model.

More specifically, we keep the LHS of equation (5) ($R_{\rm ion}$ term) same and vary the integrand of RHS integrals, namely $\mathbb{R}_{\rm rec,\,V}$ and $N_{\rm HI,\,V}$, to create the following models.

(i) Full-NH-Full-Rrec (FNH-FRrec): $N_{\rm H} + R_{\rm rec}$.
(ii) Full-NH-Partial-Rrec (FNH-PRrec): $N_{\rm H} + x_{\rm H\,II}\,R_{\rm rec}$.
(iii) Full-NH-No-Rrec (FNH-NRrec): $N_{\rm H}$.
(iv) Partial-NH-Full-Rrec (PNH-FRrec): $(1-x_{\rm H\,II})\,N_{\rm H} + R_{\rm rec}$.
(v) Partial-NH-Partial-Rrec (PNH-PRrec): $(1-x_{\rm H\,II})\,N_{\rm H} + x_{\rm H\,II}\,R_{\rm rec}$.
(vi) Partial-NH-No-Rrec (PNH-NRrec): $(1-x_{\rm H\,II})\,N_{\rm H}$.
(vii) No-NH-Full-Rrec (NNH-FRrec): $R_{\rm rec}$.
(viii) No-NH-Partial-Rrec (NNH-PRrec): $x_{\rm H\,II}\,R_{\rm rec}$.

The time-integrated model is represented here by ionization condition of PNH-PRrec whereas the instantaneous model uses that of NNH-FRrec. The others are variants on these as summarized in Table 1. To illustrate the differences in the 21 cm power spectrum, we use the same density field and halo catalogues generated within a simulation run of a box size 75 Mpc and $N = 140^3$. We have shown previously (fig. 8 in Hassan et al. 2016) that the numerical volume convergence of our simulated 21 cm power spectrum is excellent at all redshifts down to a box size of 75 Mpc; hence, we expect the same 21 cm power spectrum for larger simulation volumes. The reionization histories produced by these models vary, so as before we choose to make our 21 cm power spectrum comparison at a fixed neutral fraction.

Fig. 6 shows the 21 cm power spectrum produced by different models at 50 per cent neutral fraction as explained above. First, we

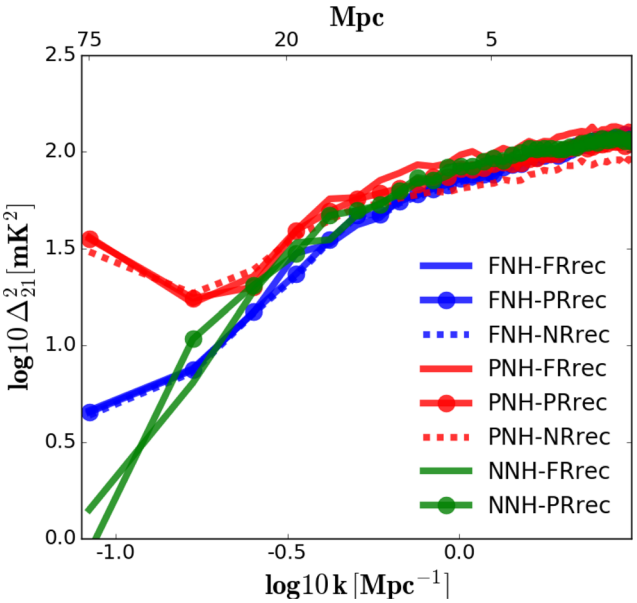

Figure 6. The 21 cm power spectrum comparison for different physical assumptions at $x_{\rm H\,I} \sim 50$ per cent. Different colours represent different ways to treat the local fluctuations in the neutral hydrogen density while different line styles correspond to different recombination terms, as explained by the legend and text. It is evident that the recombinations are subdominant in determining the large-scale 21 cm power spectrum. It is also shown that models (red) that track the neutral fraction from partially ionized regions yield a very high 21 cm power spectrum on large scales.

see that all these variants result in virtually the same 21 cm power spectrum on small scales. This reiterates our previous finding in Hassan et al. (2016) that using a non-linear ionization rate $R_{\rm ion}$ boosts the 21 cm power spectrum by a similar amount regardless of whether one accounts for recombinations or not, and further shows that accounting for the neutral hydrogen atoms does not alter this conclusion.

More significant differences are evident at large scales for the 21 cm power spectrum. Models starting with full $N_{\rm H}$ (FNH-) produce 21 cm power spectra with the same shape and amplitude on all scales (i.e. all the blue lines overlap), and likewise for models with partial or no $N_{\rm H}$ (PNH- and NNH-). This demonstrates that recombinations are subdominant for determining the large-scale 21 cm power spectrum. It is clear that the NH term plays a major role in boosting/suppressing the large-scale 21 cm power spectrum. This means that semi-numerical models must carefully account for the local number density of neutral hydrogen for a proper prediction of the expected signal.

From Fig. 6, we see the clear trend that models that do not account for the existing neutral hydrogen atoms (NNH- models such as the instantaneous model) have lower 21 cm power spectrum on large scales. Furthermore, models that use the total number of hydrogen atoms (FNH- models) at each time-step regardless of the ionization fraction show 21 cm power spectrum that is slightly higher on large scales as compared to NNH- models. This is due to the presence of weak H I fluctuations by following only the density field ($N_{\rm H\,I} \sim \Delta$). However, models that use the ionization history of cells to track the neutral hydrogen atoms from partially ionized regions (PNH- models such as the time-integrated model) show a very high 21 cm power spectrum on large scales as opposed to the NNH- and FNH- models. This comes from the fact that the PNH- models account for strong H I fluctuations by following the density field ($N_{\rm H\,I} \sim \Delta$) and ionization field ($N_{\rm H\,I} \sim x_{\rm H\,II}$) both. This shows that, at given neutral fraction, the large-scale 21 cm power spectrum is highly influenced by the way in which we account for the fluctuations in the local neutral hydrogen density.

In the next section, we will discuss the calibration of the time-integrated model against various EoR key observables and test how well the ongoing/upcoming 21 cm observations will further constrain our free parameters.

5 MODEL CALIBRATION

We now focus on our favoured time-integrated reionization model, which includes all our new physics implementations. Previously, the parametrization of $R_{\rm ion}$ (equation 2) was obtained from our small-volume high-resolution radiative transfer hydrodynamic simulation (6/256-RT). However, the small volume of this simulation makes it subject to uncertainties since, as we saw in Fig. 1, the ionization history of this simulation is significantly delayed by its small volume. Here, we adopt a more general form for $R_{\rm ion}$, and determine whether existing EoR measurements can calibrate our source model, and thereby provide constraints on the nature of reionizing sources.

To this end, we here consider a more generalized model with the following three free parameters.

(i) $f_{\rm esc}$ is the volume-averaged photon escape fraction.
(ii) $A_{\rm ion}$ is the ionizing emissivity amplitude, which scales the amount of ionizing emissivity ($R_{\rm ion}$) equally across the halo mass range at a given redshift.
(iii) $C_{\rm ion}$ is ionizing emissivity–halo mass power dependence, which quantifies the $R_{\rm ion}$–$M_{\rm h}$ slope.

We will constrain these three parameters against various EoR observations and compare with the values found from fitting to the 6/256-RT simulation, using a Bayesian MCMC approach. We choose these parameters to explore since they are most closely related to the emission characteristics of the source population. We ignore $B_{\rm ion}$ that is related to how photoionization suppresses low-mass galaxy growth, and $D_{\rm ion}$ because it is not physically obvious why ionization rate of a given halo should have a strong redshift dependence. While it would be better to simply let all these parameters vary, even doing an MCMC over this 3D space is already computationally challenging since it requires doing full runs for each sampling, and increasing the dimensionality quickly makes the computational requirements intractable.

Recall that the time-integrated model identifies the ionized regions using a time-dependent ionization condition (equation 5), which tracks the exact recombinations and neutral hydrogen atoms by following the reionization history. The reionization history, in this model, is numerically well converged for $\Delta z \leq 0.125$. With these requirements, the model becomes more computationally expensive to run, but nevertheless is feasible for independent large-volume runs. However, sampling the full MCMC space requires at least $\sim 10^6$ simulation realizations, which becomes infeasible. Hence, we pre-compute a grid of models spanning the full prior space, and then do a trilinear interpolation to obtain the observables for any given parameter combination. This sacrifices some accuracy but makes the computation feasible.

We note that Greig & Mesinger (2015) developed an analysis pipeline, 21CMMC, that directly links their semi-numerical model 21CMFAST (Mesinger, Furlanetto & Cen 2011) to a Bayesian routine COSMOHAMMAR (Akeret et al. 2013) to constraining their free parameters. However, their ionization condition did not include

recombinations through the time integral method that they developed in Sobacchi & Mesinger (2014), and instead used a standard efficiency parameter (ζ) approach. Along with lower resolution of ∼2 Mpc, these simplifications enabled them to run their semi-numerical model fully within an MCMC scheme.

5.1 Parameter estimation pipeline

We choose a cell size of 0.375 h^{-1} Mpc and a box size $L = 75$ Mpc, giving $N = 140$ cells per side. We pre-compute a grid of $25 \times 25 \times 25$ runs outputting the predicted observables for our models, uniformly sampling our selected prior range for our parameters of ($f_{\rm esc}$, log10($A_{\rm ion}$), $C_{\rm ion}$) = [(0,1),(37,44),(−1,2)]. This gives a total of 15 625 simulation independent realizations, which we interpolate inside the MCMC search process.

We have tested our parameter constraints using two different Bayesian inference tools: MULTINEST (Feroz, Hobson & Bridges 2009) and EMCEE (Foreman-Mackey et al. 2013). We have found the same parameter estimates using these two different codes, and hence, our presented parameters estimation here appear to be robust to variations in the algorithm used.

We here present the results obtained by using the EMCEE python package. We use 100 random walkers initialized around the maximum likelihood. For each walker, we sample 10 000 chains from the likelihood after 500 initial burn-in chains to achieve convergence. This makes a total of 1000 000 samples that is sufficient to explore the whole parameter space.

5.2 EoR key observable constraints

We constrain our three free parameters to the following observations.

(i) The dust-corrected star formation rate (SFR) density integrated down to $M_{\rm AB} = -17$ by Bouwens et al. (2015) at the following redshifts:

(a) $z \sim 6$: log10(SFR) [$M_\odot$ Mpc^{-3} yr^{-1}] $= -1.55 \pm 0.06$.
(b) $z \sim 7$: log10(SFR) [$M_\odot$ Mpc^{-3} yr^{-1}] $= -1.69 \pm 0.06$.
(c) $z \sim 8$: log10(SFR) [$M_\odot$ Mpc^{-3} yr^{-1}] $= -2.08 \pm 0.07$.

(ii) The Planck Collaboration XLVII (2016b) integrated optical depth to Thomson scattering: $\tau = 0.058 \pm 0.012$.

(iii) The Becker & Bolton (2013) ionizing emissivity density measurements from Lyα data at $z = 4.75$: $\dot{N}_{\rm ion}$ [10^{51} photons s^{-1} Mpc^{-3}] $= -0.014^{+0.454}_{-0.355}$.

We will first examine how our free parameters are constrained individually by each observation, and then we will examine the combined constraints.

5.2.1 *The Bouwens et al. (2015) SFR constraints*

Unlike other semi-numerical models that rely on the efficiency parameter ζ, our model allows a direct comparison to the SFR measurements by using a parametrization for $R_{\rm ion}$ that is directly relatable to SFR. For a consistent comparison with Bouwens et al. (2015) measurements, we convert the $R_{\rm ion}$ back to SFR using equation 2 in Finlator, Davé & Özel (2011) that is based on Schaerer (2003) models, and add up all SFR from haloes brighter than $M_{\rm AB} = -17$ at $z = 6, 7, 8$. To compute the corresponding $M_{\rm AB}$, we use the linear relation provided in Kennicutt (1998) that converts the SFR to luminosity L_ν over the wavelength range 1500–2800 Å.

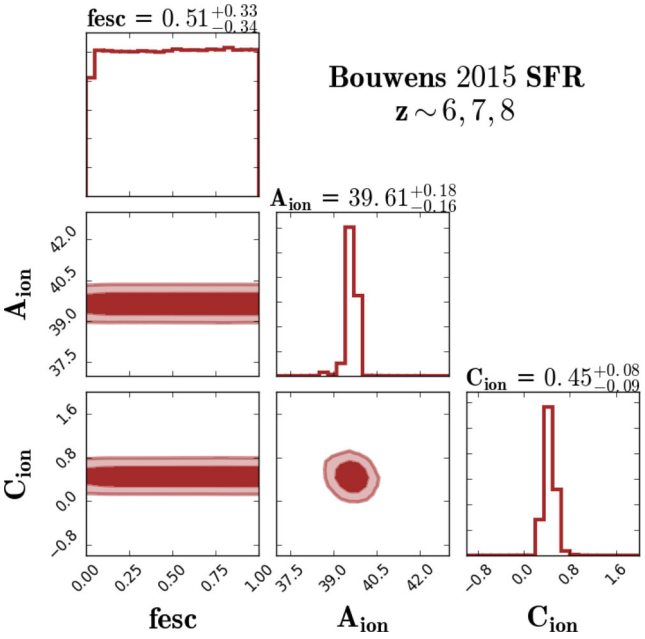

Figure 7. Bouwens et al. (2015) SFR observation, at $z = 6, 7, 8$ combined together, constraints on our model parameters. Values on top of the 1D PDFs' diagonal represent the best-fitting parameters with 1σ (14th and 84th percentiles). Dark and light shaded regions correspond to 1σ and 2σ levels, respectively. The Bouwens et al. (2015) SFR observations provide tight constraints on the $A_{\rm ion}$ and $C_{\rm ion}$ in the selected prior range. As expected, the SFR measurements do not constrain the $f_{\rm esc}$.

Fig. 7 shows the posterior distribution of our parameters as constrained solely by Bouwens et al. (2015) integrated SFR observations at $z = 6, 7, 8$ (taken together). This provides somewhat tighter constraints than fitting to a single redshift of SFR measurement, although constraining to a single redshift yields similar results, which indicates that the weak redshift evolution in the SFR measurements is adequately reproduced by our model for $R_{\rm ion}$.

The Bouwens et al. (2015) SFR observations provide tight constraints on the $A_{\rm ion}$ and $C_{\rm ion}$ as seen in Fig. 7. while poorly constraining $f_{\rm esc}$. The latter is expected because the $f_{\rm esc}$ is set by the recombinations in the interstellar medium (ISM) while the SFR depends on the halo mass and redshift.

The value of $C_{\rm ion} = 0.45$ agrees within the 1σ level with what was previously found from fitting our hydrodynamical simulations, which yielded $C_{\rm ion} = 0.41$ (Hassan et al. 2016). This means that our large-volume semi-numerical model is compatible with the same slope of the $R_{\rm ion}$–$M_{\rm h}$ relation predicted by the small-volume 6/256-RT simulation to match the Bouwens et al. (2015) SFR observation, thereby nicely corroborating the direct simulation results.

However, the differences are more significant in the $A_{\rm ion}$ posterior distribution. We see that the $A_{\rm ion}$ best-fitting value of $10^{40.03}$ predicted by 6/256-RT simulation overestimates by ∼50 per cent the value of $A_{\rm ion} = 10^{39.61}$ favoured by our SIMFAST21 MCMC fit using only the Bouwens et al. (2015) SFR constraints. This represents somewhat poor concordance at only a 3σ level. This discrepancy arises due to the small box size (=6 h^{-1} Mpc) of 6/256-RT simulation that does not capture the large-scale fluctuations and massive dark matter haloes that contribute significantly to the reionization photon budget. Hence, the 6/256-RT simulation requires larger $A_{\rm ion}$ to compensate for these limitations. This effect is also seen in Fig. 1 when comparing the reionization histories of the 6/256-RT simulation with our large-volume semi-numerical simulations that tend to

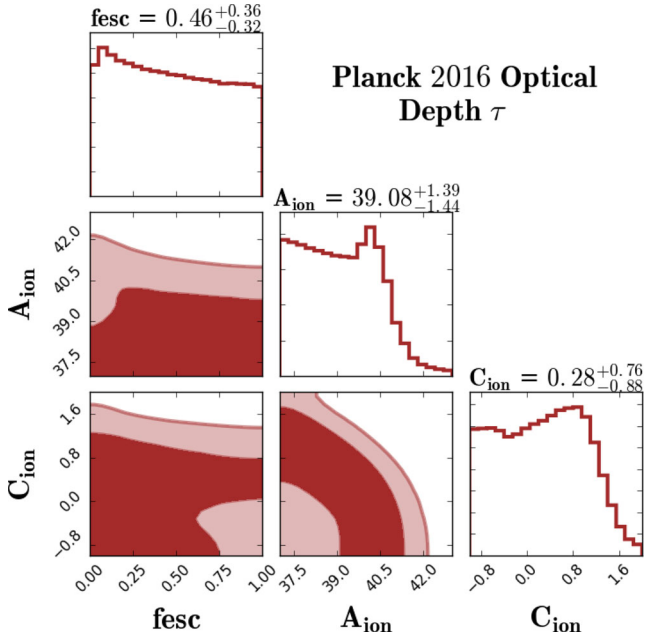

Figure 8. Planck Collaboration XLVII (2016b) optical depth constraints on our model parameters. Values on top of the 1D PDFs' diagonal represent the best-fitting parameters with 1σ (14th and 84th percentiles). Dark and light shaded regions correspond to 1σ and 2σ levels, respectively. The Planck Collaboration XLVII (2016b) τ provides poor constraints on all parameters while there is a slight tendency towards lower $f_{\rm esc}$ values. The Planck Collaboration XLVII (2016b) τ prefers models with low $A_{\rm ion}$ and $C_{\rm ion}$ values for the chosen prior range as compared to values implied by the Bouwens et al. (2015) SFR observations.

reionize the Universe much earlier at a fixed optical depth due to the presence of those massive haloes and large-scale fluctuations.

Overall, utilizing only the integrated SFR observations already gives interesting constraints on the slope and amplitude of the ionizing photon output as a function of halo mass. However, there are no useful constraints on the escape fraction.

5.2.2 *The Planck Collaboration XLVII (2016b) optical depth constraints*

Fig. 8 shows the parameters constrained to match solely the Planck Collaboration XLVII (2016b) data. This shows that the Thomson optical depth data alone provides fairly poor constraints on any of the parameters. There is a slight tendency to favour lower $f_{\rm esc}$ values, as also found by Greig & Mesinger (2017, see their fig. 4), but in general all values from zero to one are still allowed.

The main reason for the lack of sensitivity to $f_{\rm esc}$ is shown in Fig. 9, and essentially arises from the still large errors on τ. Fig. 9 shows the volume-weighted global neutral fraction evolution for fixed values of $A_{\rm ion}$ and $C_{\rm ion}$, and shows that $f_{\rm esc} = 20 \rightarrow 80$ per cent gives rise to $\tau = 0.058 \rightarrow 0.071$, which is still essentially within the 1σ uncertainty on the measurement of $\tau = 0.058 \pm 0.012$. Hence, much smaller error bars on τ are required to provide better constraints on $f_{\rm esc}$.

$A_{\rm ion}$ and $C_{\rm ion}$ are also not well constrained by the Thomson optical depth data alone, though there is some tendency to favour small values of $A_{\rm ion}$ and $C_{\rm ion}$. None the less, the uncertainties are large, and the values favoured from the SFR constraints alone are within the 1σ uncertainties of these predictions, as are the values found directly from the hydrodynamic simulations.

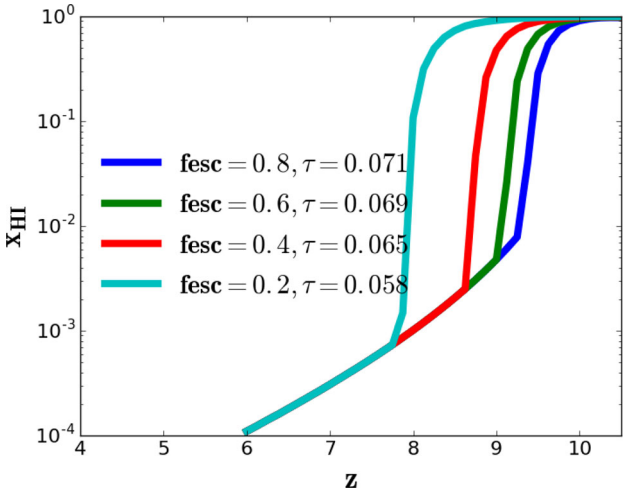

Figure 9. The reionization history for our time-integrated model with different $f_{\rm esc}$ values while fixing the $A_{\rm ion}$ and $C_{\rm ion}$ to values implied by the recent Planck Collaboration XLVII (2016b) τ measurements. This clearly shows that the current Planck Collaboration XLVII (2016b) optical depth $\tau = 0.058 \pm 0.012$ does not provide tight $f_{\rm esc}$ constraints for models with rapid reionization scenarios as the case with our time-integrated EoR model.

In summary, the Thomson optical depth as measured by *Planck* alone does not provide strong constraints on any of our parameters. It is clear that reducing uncertainties and/or including other data will be required in order to meaningfully constrain the sources driving reionization.

5.2.3 *The Becker & Bolton (2013) ionizing emissivity constraints*

The integrated emissivity of ionizing photons $\dot{N}_{\rm ion}$ quantifies the total ionization rate density from all ionizing sources that escape galaxies to fill the IGM. Mathematically, $\dot{N}_{\rm ion} = \sum f_{\rm esc} R_{\rm ion}$ divided by the simulation's comoving volume. To compare with Becker & Bolton (2013) $\dot{N}_{\rm ion}$ measurements, we add up $f_{\rm esc} R_{\rm ion}$ from all haloes and divide by the simulation comoving volume at $z = 4.75$. As with the SFR data, our model permits a direct comparison with the $\dot{N}_{\rm ion}$ data since we use a parametrization for $R_{\rm ion}$ rather than a single efficiency parameter.

Fig. 10 shows the posteriors for our three free parameters constrained only to match the Becker & Bolton (2013) $\dot{N}_{\rm ion}$ data. As with the SFR and τ constraints, the $f_{\rm esc}$ is unconstrained by these data. Similar to Planck Collaboration XLVII (2016b) τ constraints, we find that models with high $A_{\rm ion}$ and $C_{\rm ion}$ values are disfavoured by Becker & Bolton (2013) $\dot{N}_{\rm ion}$ measurements, but again this is within 1σ of the SFR-only constraints.

The slight tendency of $\dot{N}_{\rm ion}$ data towards negative values of $C_{\rm ion}$ (negative slope of $R_{\rm ion}$–$M_{\rm h}$ relation) favours small haloes being the dominant ionizing photon source contributor to match the post-reionization measurements. In contrast, SFR and τ data prefer the positive side of $C_{\rm ion}$ values, implying that massive haloes are more important during the reionization. This shows that reionization requires more ionizing photons while matching post-reionization data requires fewer ionizing photons. This stands as one of the theoretical challenges for the EoR models as it is not easy to match simultaneously observational constraints during and after reionizations. $\dot{N}_{\rm ion}$ measurements at higher redshifts ($z \sim 6$, 7) would be very useful to see if this tension extends into the overlapping redshift regime (Keating et al. 2014).

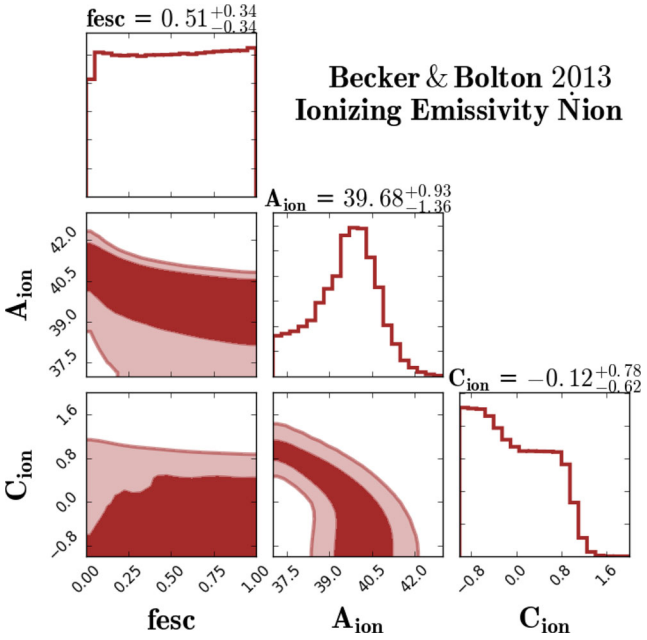

Figure 10. The Becker & Bolton (2013) ionizing emissivity ($z = 4.75$) constraints on our model parameters. Values on top of the 1D PDFs' diagonal represent the best-fitting parameters with 1σ (14th and 84th percentiles). Dark and light shaded regions correspond to 1σ and 2σ levels, respectively. Similar to previous constraints, the $f_{\rm esc}$ is poorly constrained and similar to τ constraints, the data prefer lower $f_{\rm esc}$ values. The $\dot{N}_{\rm ion}$ data also prefer models with negative $C_{\rm ion}$ in our selected prior range. This shows that matching to post-reionization data requires fewer ionizing photons and prefers models with dominant contributions from small dark matter haloes.

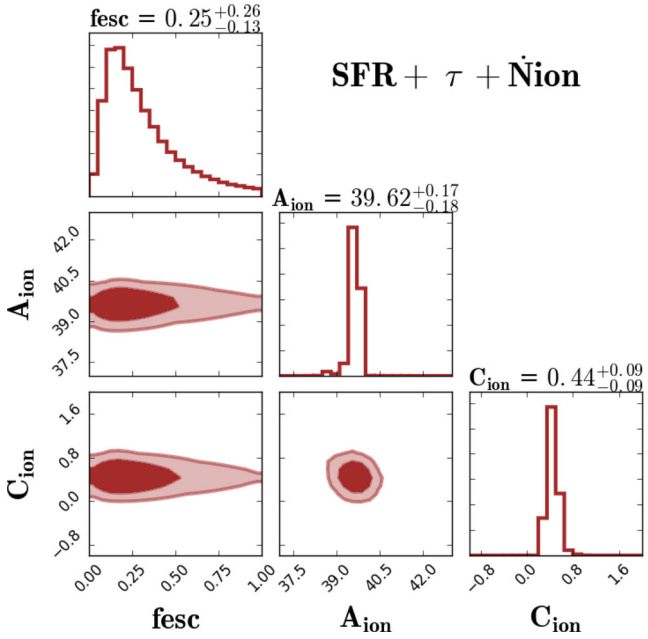

Figure 11. Combined constraints from SFR, τ and $\dot{N}_{\rm ion}$. Values on top of the 1D PDFs' diagonal represent the best-fitting parameters with 1σ (14th and 84th percentiles). Dark and light shaded regions correspond to 1σ and 2σ levels, respectively. Combining all these observations results in a tighter $f_{\rm esc}$ constraint while $A_{\rm ion}$ and $C_{\rm ion}$ still follow Bouwens et al. (2015) SFR constraints.

5.2.4 *Combined SFR + τ + $\dot{N}_{\rm ion}$ constraints*

To obtain the strongest constraints given the observations we consider, we now combine our three key EoR constraints: the Bouwens et al. (2015) SFR observations, Planck Collaboration XLVII (2016b) optical depth measurements and the Becker & Bolton (2013) $\dot{N}_{\rm ion}$ data. This represents the best available constraints we can make given current data, and serves to provide our base model from which we will do forecasting for 21 cm experiments.

Fig. 11 shows the parameter estimates as fitted to the combined sample of these EoR observations. We see that the $A_{\rm ion}$ and $C_{\rm ion}$ are tightly constrained, which as Fig. 7 showed is driven by the Bouwens et al. (2015) SFR constraints, as the other observations did not provide very tight constraints on these parameters.

The more interesting difference is in $f_{\rm esc}$, where the combined constraints now definitely prefer lower $f_{\rm esc}$ values, with best-fitting value of $0.25^{+0.26}_{-0.13}$. This is still a rather wide range, and the posterior ellipses show that even very low escape fractions are not ruled out at more than a $\sim 1\sigma$ level, and very high $f_{\rm esc}$ values are only disfavoured at $\lesssim 2\sigma$. This tendency was hinted at from matching to Becker & Bolton (2013) $\dot{N}_{\rm ion}$ and Planck Collaboration XLVII (2016b) optical depth individually. This result indicates that our previous findings of $f_{\rm esc} = 0.04$–0.06 in Hassan et al. (2016) are clearly possible for models with higher $A_{\rm ion}$ and $C_{\rm ion}$ values within their derived 1σ level. A summary of the individual and combined constraints is provided in Table 2.

This shows that current observations can already constrain the basic power-law parameters of the ionizing photon output versus halo mass, but constraints on $f_{\rm esc}$ are still somewhat elusive. Note that we are also assuming a constant $f_{\rm esc}$ for all galaxies, while there may be some mass and/or redshift dependence; however, with even a single parameter already being poorly constrained, it is unlikely that adding more parameters will allow tighter constraints.

6 21 CM FORECASTING AND EXPERIMENT SENSITIVITIES

The ultimate goal is to add the 21 cm observations to these existing data (or future improved versions thereof), in order to ascertain how well we can understand the sources of reionization. To do so, we adopt a forecasting approach by which we use expected uncertainties from future 21 cm power spectrum measurements in concert with these existing data and ascertain how much improvement the 21 cm data will provide in the precision with which our parameters are constrained. We will assume a base model that is the best fit to our current constraints as listed in Table 2.

We focus our analysis on LOFAR, HERA and SKA1-Low. For each experiment, we first compute the thermal noise power spectrum that dominates the errors in measuring the 21 cm signal. We then add more uncertainties from the sample variance, while neglecting the shot noise since it has been shown to have a minimal effect at the relevant scales ($k < 2\,h$ Mpc^{-1}) for these telescope sensitivities (Pober et al. 2013). We obtain these uncertainties using the 21CMSENSE package,[4] and refer to Parsons et al. (2012) for the full mathematical derivation of the radio interferometer sensitivities, and to Pober et al. (2013, 2014) for more details on observation strategies and foreground removal models. We briefly highlight the basic equations and concepts used in 21CMSENSE to obtain the 21 cm power spectrum error from a specific array configuration.

[4] https://github.com/jpober/21cmSense

Table 2. Summary of our parameter estimations from individual and combined set of observations, as well as from matching to the 6/256-RT simulations.

EoR constraint	$f_{\rm esc}$	$\log10(A_{\rm ion})$	$C_{\rm ion}$
Bouwens et al. (2015) SFR all at $z = 8, 7, 6$	$0.51^{+0.33}_{-0.34}$	$39.61^{+0.18}_{-0.16}$	$0.45^{+0.08}_{-0.09}$
Planck Collaboration XLVII (2016b) optical depth τ	$0.46^{+0.36}_{-0.32}$	$39.08^{+1.39}_{-1.44}$	$0.28^{+0.76}_{-0.88}$
Becker & Bolton (2013) ionizing emissivity $N_{\rm ion}$ at $z = 4.75$	$0.51^{+0.34}_{-0.34}$	$39.68^{+0.93}_{-1.36}$	$-0.12^{+0.78}_{-0.62}$
ALL = SFR+τ+$N_{\rm ion}$	$0.25^{+0.26}_{-0.13}$	$39.62^{+0.17}_{-0.18}$	$0.44^{+0.09}_{-0.09}$
Values obtained from fitting to 6/256-RT	–	40.03	0.41

The dimensionless power spectrum of the thermal noise (Parsons et al. 2012; Pober et al. 2013, 2014) can be obtained using

$$\Delta^2_N(k) \approx X^2 Y \frac{k^3}{2\pi^2} \frac{\Omega}{2t} T^2_{\rm sys}\,, \tag{7}$$

where X^2Y is a conversion factor from angle and frequency units to comoving cosmological distances, Ω is the primary beam field of view, t is the integration time and $T_{\rm sys}$ is the system temperature (sky+receiver). It is then straightforward to add the sample variance to the thermal noise to obtain the total error (Pober et al. 2013) as follows:

$$\delta\Delta^2(k) = \left(\sum \frac{1}{\left(\Delta^2_N(k) + \Delta^2_{21}(k)\right)^2} \right)^{-\frac{1}{2}}, \tag{8}$$

where Δ^2_{21} is the 21 cm power spectrum and the summation runs over all measured independent k-modes.

We construct these experiments as follows:

(i) *LOFAR*. We use the Netherlands 48 High-Band Antennas (HBA) with positions listed in van Haarlem et al. (2013) following Pober et al. (2014). Each antenna has a diameter of 30.75 m that results in a total collecting area of 35 762 m^2 for the 48 HBA station. The receiver temperature $T_{\rm rcvr}$ is set to 140 000 mK as suggested by Jensen et al. (2013) and Greig & Mesinger (2015).

(ii) *HERA*. We consider the final design of 331 hexagonally packed 14 m antennas (Beardsley et al. 2015; Ewall-Wice et al. 2016). With this configuration, the total collecting area becomes 50 953 m^2. We assume a 100 000 mK receiver temperature $T_{\rm rcvr}$, similar to previous works by Pober et al. (2014) and Greig & Mesinger (2015).

(iii) *SKA-LOW1*. We model SKA1-Low following the SKA1 System Baseline Design document by Dewdney (2013) in which the proposed array consists of 911 antennas in total. These antennas are distributed randomly to form a compact core using 866 dishes surrounded by the remaining 45 dishes along spiral arms. The 866 core antennas provide the vast majority of the sensitivity, and hence our SKA model ignores those 45 spiral arm stations (Pober et al. 2014; Greig & Mesinger 2015). Each station of 866 antennas has a diameter of 35 m that makes a total collecting area of 833 189 m^2. The receiver noise here is determined by $T_{\rm rcvr} = 0.1\ T_{\rm sky} + 40$ K, where the sky temperature is modelled using $T_{\rm sky} = 60\lambda^{2.55}$.

For a consistent comparison, we choose to operate these three array designs in a drift-scanning mode for 6 observing hours per day for 180 d at 8 MHz bandwidth. We consider Pober et al. (2014) moderate foreground removal model where the foreground wedge extends 0.1 h Mpc^{-1} beyond horizon limit.

6.1 Including the 21 cm data

We combine three different redshifts of 21 cm power spectrum observations, namely $z = 9.0, 8.5, 8.0$, which provides tighter constraints than considering any single-epoch observations. With multiple redshift 21 cm observations, one accounts simultaneously for the variation in redshift (density field) and neutral fraction (ionization field) evolution, which are the main components in determining the 21 cm fluctuations. Given the rapid reionization behaviour of the time-integrated model as shown in Fig. 1, our selected redshifts ($z = 9, 8.5, 8$) correspond to a wide range of neutral fractions that account for different reionization epochs such as the initial bubble growth and the bubble overlap phase. We next construct the likelihood from these observations by simply adding up their individual χ^2. We limit our analysis to a wide k-range of 0.15–1.0 Mpc^{-1}, consistent with Greig & Mesinger (2015). From this k-range, we select 10 bins of the power spectrum that is sufficient to capture the fluctuations for a given 21 cm power spectrum.

We use the well-calibrated time-integrated model with parameters derived from fitting to our combined set of EoR observations as discussed in Section 5.2.4 and shown in Fig. 11. Specifically, we use the following parameters: $(f_{\rm esc}, \log10(A_{\rm ion}), C_{\rm ion}) = (0.24, 39.63, 0.43)$, consistent with the 1σ level of constraints by our combined set of EoR observations. We then use these parameters to create our mock observations with a large box size of $L = 300$ Mpc and $N = 560$ per side that results in a resolution of 0.375 h^{-1} Mpc. We determine the error in measuring the 21 cm power spectra for our mock observation by using the telescope sensitivity code 21CMSENSE for each specific array experiment at our chosen redshifts as described above. We use the same pipeline discussed in Section 5.1 to sample the 21 cm power spectrum space, except now we include the 21 cm mock observation power spectra among the pre-computed runs to study how well the MCMC technique may recover the input model parameters.

Fig. 12 shows our 21 cm mock observations at several redshifts. The shaded area corresponds to the error in measuring the 21 cm power spectrum for our large mock observations using the 21CMSENSE package. LOFAR (green shading), operating currently, will be able to constrain only the largest scales considered here, while HERA (blue), under construction now, will be further sensitive to intermediate scales, while the future SKA1-Low (red) will provide tight constraints into the sub-Mpc scale regime owing to its wider baselines and hence better resolution. These uncertainties depend mainly on our telescope configurations as described above. Hence, the main improvement as these facilities develop will be to better constrain the 21 cm power spectrum towards smaller scales, and each generation will provide significant gains in this.

6.2 21 cm MCMC

We now ask how well the 21 cm data can constrain our free parameters. First, we consider the 21 cm power spectrum data as shown in Fig. 13 by itself, to see how tightly our parameters can be constrained by such observations alone. Then we add the 21 cm data to our other existing observational constraints. In each case,

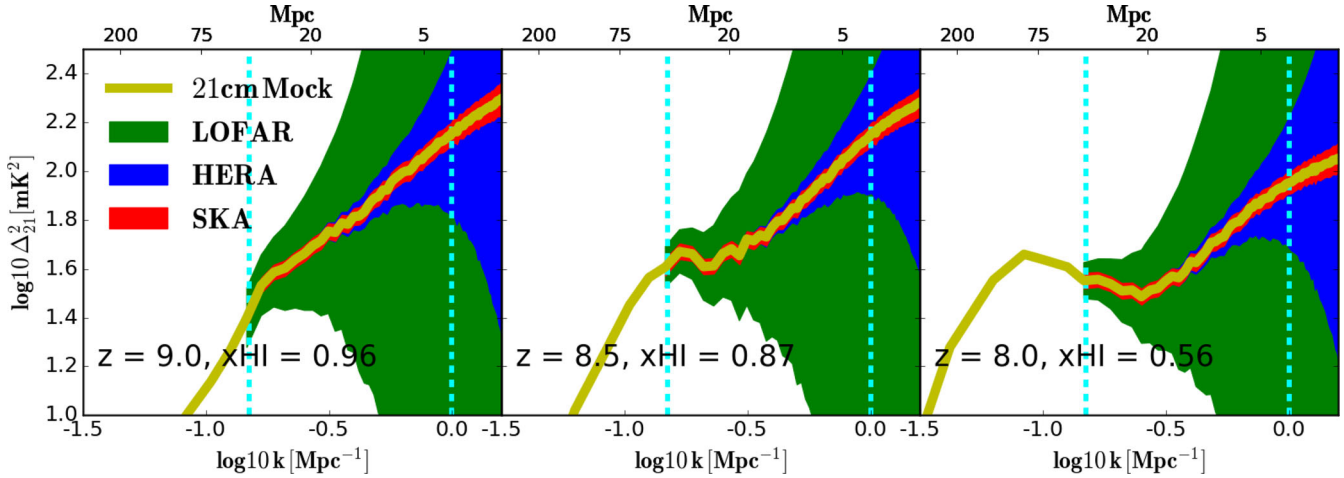

Figure 12. Three redshift mock 21 cm EoR observations using the well-calibrated time-integrated EoR model with parameters ($f_{\rm esc}$, log10($A_{\rm ion}$), $C_{\rm ion}$) = (0.24, 39.63, 0.43). Yellow solid line represents the 21 cm power spectrum from the large-box mock observation ($L = 300/N = 560$). Shaded area shows the error bars obtained using 21CMSENSE package for our constructed EoR arrays: SKA (red), HERA (blue), LOFAR (green). Redshifts and neutral fractions for 21 cm mock observations are quoted in each panel. Vertical cyan dashed lines show our chosen k-range to perform the 21 cm MCMC.

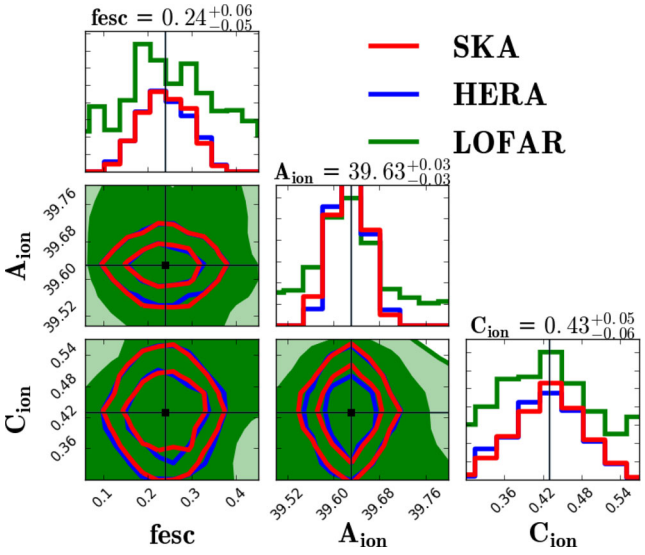

Figure 13. 21 cm power spectrum constraints on our three EoR parameters from several redshift ($z = 9.0, 8.5, 8.0$) mock observations. SKA, HERA and LOFAR constraints are shown by red, blue and green contours, respectively. Values on top of the 1D PDFs represent the best-fitting parameters as implied by the SKA mock observations while black square points correspond to the input mock observation parameters: ($f_{\rm esc}$, log10($A_{\rm ion}$), $C_{\rm ion}$) = (0.24, 39.63, 0.43). The MCMC technique is able to recover the input model parameters. It is evident that the future 21 cm observations can tightly constrain our model parameters for experiments with small and intermediate levels of uncertainty in detecting the expected signal such as SKA and HERA, respectively.

we use our MCMC framework to determine our best-fitting values of our free parameters and their uncertainties using the entire data set, for the case of each telescope facility. This provides forecasting for how much improvement can be expected from future 21 cm observations.

Fig. 13 shows the 1D Probability Density Functions (PDFs) and 2D contours of our three parameters from the combined redshifts ($z = 9.0, 8.5, 8.0$) of 21 cm mock observations by our three selected EoR experiments. To begin, we see that our MCMC search well recovers the best-fitting input model (mock observation) parameters (black square points). This is to be expected, since the same input model was used to generate the 21 cm data. The improvement to be noted here is the reduction of the uncertainties on these parameters relative to the previous case without 21 cm data.

For LOFAR (green shaded area), we see that the 21 cm observations do not provide tight constraints due to large uncertainties as seen in Fig. 12. Essentially, mildly constraining the large-scale power provides little information on the ionization sources that drive reionization.

In contrast, HERA (blue) and SKA (red) provide quite tight constraints on the free parameters. Note that the scale of the posteriors is substantially reduced relative to our previous plots in order to enhance visibility. Hence, future 21 cm data alone can already independently constrain reionizing sources, without adding in any other observations. Interestingly, there is almost no difference between the SKA and HERA constraints. This arises because the parameter constraints are predominantly driven by the larger scales, and HERA and SKA provide similar constraints on the power spectrum for scales $\gtrsim$2 Mpc.

Comparing the 21 cm constraints with constraints obtained from combining several EoR key observables, we find that constraining to 21 cm observations yields smaller parameter errors. This can be clearly seen when comparing the 1σ level of $f_{\rm esc}$ and $A_{\rm ion}$ found by constraining to the 21 cm observations (Fig. 13) versus to the combined EoR sample (SFR, $\dot{N}_{\rm ion}$, τ) (Fig. 11). However, it is evident that the 21 cm future observation can constrain the $f_{\rm esc}$ tighter than the current EoR key observables.

In previous work by Greig & Mesinger (2015), the authors used a similar semi-numerical framework and performed similar analysis to constrain their free parameters to future 21 cm mock observations. However, they did not have the photon escape fraction as a free parameter and rather constrained their efficiency parameter ζ, from which the $f_{\rm esc}$ can be computed for various assumptions about gas fraction in stars and ionizing photon number per baryon (see their equation 2). However, we here constrain the $f_{\rm esc}$ directly without making further assumptions about the gas and baryon fractions; hence, our presented $f_{\rm esc}$ results are direct, albeit the inherent photon conservation issues in these semi-numerical models, which we will discuss later.

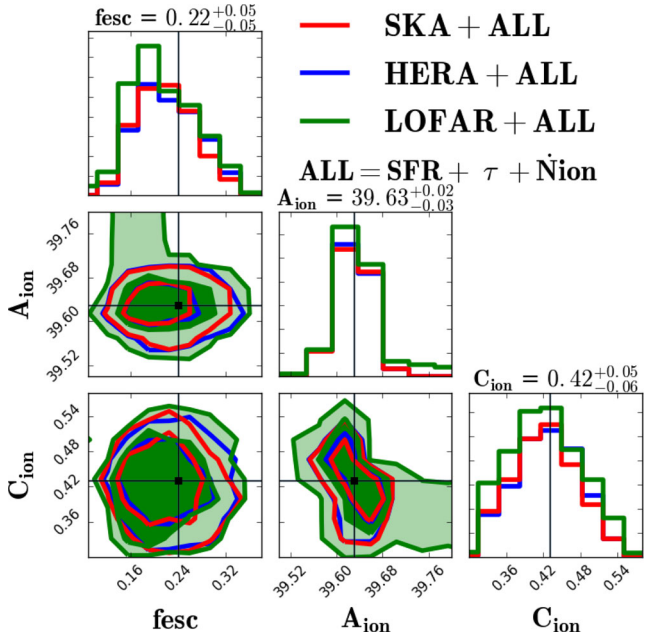

Figure 14. Parameter estimates from combining the current EoR observations with the 21 cm mock observations. The current EoR observations here are our previous combined sample: SFR, τ and $\dot{N}_{\rm ion}$ while the 21 cm mock observations are combinations of several 21 cm redshifts at $z = 9.0$, 8.5, 8.0. SKA, HERA and LOFAR constraints are shown by red, blue and green contours, respectively. Values on top of the 1D PDFs represent the best-fitting parameters as implied by the SKA mock observations while black square points correspond to the input mock observation parameters: $(f_{\rm esc}, \log10(A_{\rm ion}), C_{\rm ion}) = (0.24, 39.63, 0.43)$. It is evident that adding the current EoR observations on top of the 21 cm mock observations provides more tighter constraints even for experiments with large 21 cm power uncertainties such as the case with LOFAR.

We finally constrain our free parameters by combining the 21 cm mock observations with the current EoR key observables (SFR, $\dot{N}_{\rm ion}$, τ) as shown in Fig. 14. From this figure, we see that our three parameters are well constrained by the combined set of current EoR and 21 cm mock observations. Adding our combined EoR sample (SFR, $\dot{N}_{\rm ion}$, τ) on top of the 21 cm mock observations improves the error in estimating our free parameters, particularly for arrays with large 21 cm error bars such as LOFAR. This shows that the future 21 cm observations are important in constraining the model astrophysical parameters and complement the other existing EoR various observations. A summary of our 21 cm mock observations constraints combined with the other EoR observations is given in Table 3.

Our 21 cm forecasting shows that the future 21 cm power spectrum observations will be crucial for providing tight constraints on various parameters related to the sources of reionization. Even by themselves, such data will provide improve constraints over what can be obtained using current observations. When combined with other observations, the constraints get quite tight, even for the difficult-to-constrain photon escape fractions $f_{\rm esc}$. The tightness of the constraints suggests that it may be possible to independently constrain variations in the escape fraction with mass or redshift; we will examine this in future work.

6.3 Photon conservation

To make use of our $f_{\rm esc}$ constraints, we here test the photon conservation problem in our semi-numerical model. Previous semi-numerical models, based on the excursion-set formalism, have pointed out a violation in the photon number conservation. In Zahn et al. (2007), the authors found that their semi-numerical model loses about 20 per cent photons. They have argued that this photon loss arises from ionized bubbles overlapping, which they compensated by boosting the efficiency parameter ζ. More recent work by Paranjape, Choudhury & Padmanabhan (2016) has developed a Monte Carlo model of bubble growth to resolve the photon conservation problem in their semi-numerical model. Although their bubble growth model did not resolve the problem completely, nevertheless improvements have been achieved and they have demonstrated that the problem comes from the fact that the excursion-set-based models use the average mass of the bubbles rather than tracking the actual mass of sources and bubble local density fluctuations.

However, there are two methods to flag the spherical regions as ionized in the excursion-set formalism. The first is to flag the whole cells in the bubble (whole flagging) whereas the second is to flag only the centre cell of the bubble (centre flagging). We next use these two methods to verify the photon conservation in our time-integrated EoR model. We would expect that, during time interval dt, the total number of escaped ionizing photons ($f_{\rm esc} R_{\rm ion}\,{\rm d}t$) minus the total number of recombinations ($R_{\rm rec}\,{\rm d}t$) should be equal to the number of ionizations in the neutral hydrogen atoms ($(x_{\rm H\,II}(t_{i+1}) - x_{\rm H\,II}(t_i))N_{\rm H}$). In other words, the successful photons that manage to escape from the ISM (corrected by $f_{\rm esc}$) and from high-density regions along the way (subtracted by $R_{\rm rec}\,{\rm d}t$) should be equal to the total number of neutral atoms that have been ionized during time dt. We can write the photon conservation ratio as follows:

$$\text{Photon conservation ratio} = \frac{(x_{\rm H\,II}(t_{i+1}) - x_{\rm H\,II}(t_i))N_{\rm H}}{(f_{\rm esc} R_{\rm ion} - R_{\rm rec})\,{\rm d}t}, \quad (9)$$

Table 3. Summary of our parameter estimations from the 21 cm mock observations and from combining the 21 cm mock observations with the current EoR observations (SFR, $\dot{N}_{\rm ion}$, τ).

	$f_{\rm esc}$	log10($A_{\rm ion}$)	$C_{\rm ion}$
21 cm mock observations			
SKA	$0.240^{+0.056}_{-0.054}$	$39.628^{+0.030}_{-0.032}$	$0.431^{+0.052}_{-0.056}$
HERA	$0.237^{+0.061}_{-0.054}$	$39.626^{+0.031}_{-0.025}$	$0.425^{+0.055}_{-0.058}$
LOFAR	$0.415^{+0.384}_{-0.239}$	$39.229^{+0.606}_{-1.117}$	$0.445^{+0.341}_{-0.274}$
21 cm mock observations + ALL (SFR, $\dot{N}_{\rm ion}$, τ)			
SKA+ALL	$0.217^{+0.052}_{-0.048}$	$39.631^{+0.024}_{-0.029}$	$0.423^{+0.053}_{-0.057}$
HERA+ALL	$0.221^{+0.058}_{-0.051}$	$39.630^{+0.029}_{-0.029}$	$0.427^{+0.053}_{-0.059}$
LOFAR+ALL	$0.206^{+0.069}_{-0.045}$	$39.634^{+0.042}_{-0.033}$	$0.421^{+0.065}_{-0.060}$

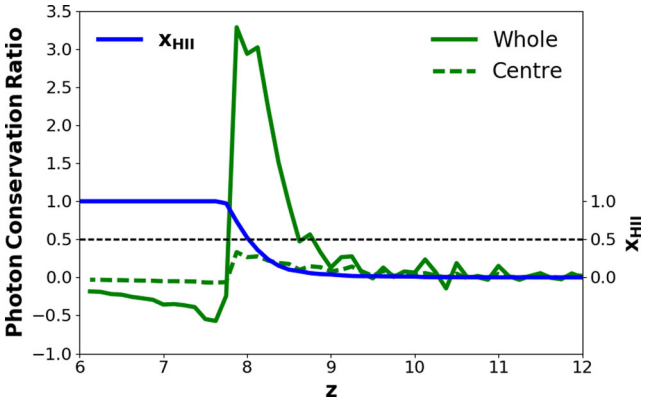

Figure 15. Photon conservation ratio from the time-integrated EoR model using whole flagging (green solid) versus centre flagging (greed dashed) scheme, with their reionization history (blue solid). Horizontal black dashed line represents the 50 per cent neutral/ionized fraction limit. Both methods violate the photon number conservations as seen by the underionization at high redshifts and overionization at the end of reionization.

where $\mathrm{d}t = t_{i+1} - t_i$. This ratio should be equal to unity for an ideal photon-conserving model. However, the ratio can be less than unity when the Universe is highly ionized. We then apply this ratio to the two methods, whole flagging versus centre flagging, to check the photon conservation problem in both. We note that centre flagging scheme requires about 20 per cent more ionizing photons to match the reionization history obtained by whole flagging method. We then adjust the $f_{\rm esc}$ in two methods to reproduce identical reionization history (identical τ) while keeping other parameters fixed.

In Fig. 15, we plot the photon conservation ratio for the two methods, whole flagging (green solid line) and centre flagging (green dashed line) with the reionization history (blue solid line). We find that the centre flagging scheme underuses photons during all reionization redshifts, even after reionization ($z < 8$), which might partly explain the need for higher $f_{\rm esc}$ with this method. The photon loss in the whole flagging scheme agrees qualitatively with centre flagging at higher redshifts when the Universe is almost neutral.

As reionization proceeds, the whole flagging starts to overuse photons and ionizes more neutral atoms than expected. The photon excess/loss in the two methods is clearly redshift dependent. In the centre flagging method, the photon loss is by a factor of ∼3,7,20 at $z = 7.75$, 9.25, 11, respectively. The whole flagging scheme shows photon loss (underusing photons) at high redshifts and photon excess (overusing photons) at the end of reionization. At high redshifts, the photon loss, in the whole flagging, is by a factor of ∼2,4,7 at $z = 8.75$, 9, 11, respectively. This shows that, at high redshifts, the photon loss, in the whole flagging method, is less by a factor of ∼2,3 as compared with centre flagging method. At $z = 8.5$ ($x_{\rm H\,I} \sim 0.9$), the whole flagging method satisfies the photon conservation condition as the ratio becomes unity, but the ratio does not converge at unity afterwards. After this point, the whole flagging scheme starts to overuse photons increasingly by a large amount till the end of reionization. We find that the photon excess is about 10 per cent at $z = 8.4$ and 70 per cent at $z = 7.75$ (end of reionization).

We note that all our EoR models adopt the whole flagging method. This shows that our constrained photon escape fractions $f_{\rm esc}$ are, in fact, overestimated by the photon excess associated with the whole flagging method. All our previous $f_{\rm esc}$ estimations can be corrected and lowered by 10 per cent up to 70 per cent depending on redshifts. The photon loss/excess evolution in redshift suggests that the $f_{\rm esc}$ might be required to change with redshift in order to preserve photon number conservation as a temporary solution.

7 CONCLUSION

We have improved our SIMFAST21 semi-numerical code for computing the EoR on large scales by incorporating a more physically motivated criterion for determining whether a region of space is ionized, as well as integrating our framework into a full MCMC parameter search framework so we can forecast how well current and future observations can constrain the physical properties of the sources driving reionization.

We have calibrated our new model to various current observations of the EoR, namely the Bouwens et al. (2015) SFR observations, the Planck Collaboration XLVII (2016b) optical depth measurements and the Becker & Bolton (2013) $\dot{N}_{\rm ion}$ data. We also compared our new EoR model to our previous EoR model in Hassan et al. (2016) in terms of their EoR history, H II bubble sizes and 21 cm power spectra. We further studied variations in the 21 cm fluctuations produced by all possible variants of our ionization conditions.

We then presented a robust MCMC analysis to constrain our generalized source model's free parameters against current EoR observations. We used the well-calibrated EoR model to predict the 21 cm power spectrum for the future EoR array experiments SKA, HERA and LOFAR. We show how the future 21 cm observations are important for complementing the existing EoR current observations in order to tightly estimate the astrophysical parameters of EoR sources.

Our key findings are as follows.

(i) The time-integrated EoR model produces very large H II bubbles as compared with the instantaneous EoR model and fewer small bubbles. This difference is clearly shown in their evolving H I maps (Fig. 2) and the ionization field (Fig. 3). This results in a larger ionization and 21 cm power spectrum on large scales by 1 to 1.2 orders of magnitude as seen in Figs 4 and 5.

(ii) By considering all possible combinations between the hydrogen atoms and recombination terms in the ionization condition, we showed that recombinations are subdominant in determining the 21 cm power spectrum particularly on large scales (Fig. 6). The 21 cm power spectrum amplitude and shape are highly sensitive to accounting for the amount of fluctuations in the neutral hydrogen density. This means semi-numerical models must carefully account for the neutral hydrogen to robustly predict the expected 21 cm signal.

(iii) The Bouwens et al. (2015) SFR observations provide tight constraints on the ionizing emissivity amplitude $A_{\rm ion}$ and the slope of the $R_{\rm ion}$–$M_{\rm h}$ relation $C_{\rm ion}$, but provide no constraint on the photon escape fraction $f_{\rm esc}$ (Fig. 7). The recent Planck Collaboration XLVII (2016b) optical depth (Fig. 8) and the Becker & Bolton (2013) $\dot{N}_{\rm ion}$ measurements (Fig. 10) poorly constrain our model parameters, while they slightly prefer models with lower values of $f_{\rm esc}$, $A_{\rm ion}$ and $C_{\rm ion}$.

(iv) Combining all of SFR, τ and $\dot{N}_{\rm ion}$ together results in tighter parameter constraints, as seen in Fig. 11. The $A_{\rm ion}$ and $C_{\rm ion}$ here follow the previous constraints by the SFR observations, but combining these measurements yields better escape fraction constraints of $f_{\rm esc} = 0.25^{+0.26}_{-0.13}$, though still not very tight. The parameters determined directly from the full hydrodynamic simulations analysed in Hassan et al. (2016) are consistent with these constraints.

(v) Using the well-calibrated time-integrated EoR model, we predict the 21 cm power spectrum at different redshifts ($z = 9$, 8.5, 8)

for several constructed radio array designs, namely SKA, HERA and LOFAR (Fig. 12). While LOFAR does not provide strong constraints except at the largest scales, future experiments will tightly constrain the 21 cm power spectrum to smaller scales that can better constrain the reionizing source population.

(vi) By adding current EoR observations (SFR, τ, $\dot{N}_{\rm ion}$) to the 21 cm mock observations, we find that all experiments recover the input model parameter accurately and the parameter errors are further improved. This illustrates how future 21 cm observations can complement and substantially improve upon existing EoR observations in order to more tightly constrain the emissivity of EoR sources and their relationship to the underlying halo population.

(vii) We find that photon conservation is suboptimal owing to the way the excursion-set formalism is generically implemented in current semi-numerical codes, including SIMFAST21. The root difficulty is that cells are treated as fully neutral or fully ionized, with no possibility of intermediate ionization levels. While some tuning could be done to minimize the problem, a robust solution likely lies in replacing the excursion-set formalism with a proper photon-conserving radiative transfer approach. We leave this for future work.

We have discussed the uncertainty associated with ionization condition in the excursion-set-based models and found that a slight change in the ionization condition could lead to a big difference in the 21 cm power spectrum particularly on large scales as seen between the time-integrated and instantaneous model. A possible approach to break such degeneracy and resolve the ionization condition uncertainty is to compare these models' 21 cm power spectra to radiative transfer simulations. For this reason, we are currently developing our own radiative transfer routine (SIMFAST21-RT) and the result will be forthcoming.

The SIMFAST21 MCMC platform developed here will be applicable for a wide range of EoR forecasting science cases. By robustly incorporating all the current observables within an MCMC framework and being able to straightforwardly incorporate new data, we are building the tools necessary to optimally connect future redshifted 21 cm power spectrum from the EoR to physical quantities associated with the population of reionizing sources. Such a framework can be used to explore and constrain exotic source populations such as mini-quasars or Population III stars, as well as to potentially extend to multi-tracer cross-correlation approaches. The most immediate hurdle will be to develop a more robust yet still fast radiative transfer method that conserves photons, so we can more reliably assess the obtained source population parameters. There is much exciting work to be done as we continue to prepare for the 21 cm EoR era.

ACKNOWLEDGEMENTS

We thank Jonathan Pober for making his 21CMSENSE sensitivity code publicly available, providing the SKA antenna coordinates, helpful discussions and comments. We acknowledge helpful discussions with Andrei Mesinger, Greig Bradley, Emanuele Sobacchi, Jonathan Zwart, Girish Kulkarni, Tirthankar Roy Choudhury and Neal Katz. We also thank Dan Foreman-Mackey for his excellent MCMC code EMCEE and visualization package CORNER (Foreman-Mackey 2016). SH is supported by the Deutscher Akademischer Austauschdienst (DAAD) Foundation. RD and SH are supported by the South African Research Chairs Initiative and the South African National Research Foundation. MGS is supported by the South African Square Kilometre Array Project and National Research Foundation. This work was also supported by NASA grant NNX12AH86G. Part of this work was conducted at the Aspen Center for Physics, which is supported by National Science Foundation grant PHY-1066293. Computations were performed at the cluster 'Baltasar-Sete-Sois', supported by the DyBHo-256667 ERC Starting Grant, and the University of the Western Cape's 'Pumbaa' cluster.

REFERENCES

Akeret J., Seehars S., Amara A., Refregier A., Csillaghy A., 2013, Astron. Comput., 2, 27
Barkana R., Loeb A., 2001, Phys. Rep., 349, 125
Bauer A., Springel V., Vogelsberger M., Genel S., Torrey P., Sijacki D., Nelson D., Hernquist L., 2015, MNRAS, 453, 3593
Beardsley A. P., Morales M. F., Lidz A., Malloy M., Sutter P. M., 2015, ApJ, 800, 128
Becker G. D., Bolton J. S., 2013, MNRAS, 436, 1023
Bouwens R. J., Illingworth G. D., Oesch P. A., Caruana J., Holwerda B., Smit R., Wilkins S., 2015, ApJ, 811, 140
Bouwens J. R. et al., 2015, ApJ, 803, 34
Choudhury T. R., Haehnelt M. G., Regan J., 2009, MNRAS, 394, 960
Davé R., Katz N., Oppenheimer B. D., Kollmeier J. A., Weinberg D. H., 2013, MNRAS, 434, 2645
Dewdney P. E., 2013, Technical report, SKA1 System Baseline Design SKA-TEL-SKO-DD-001, Rev. 1
Ewall-Wice A. et al., 2016, MNRAS, 460, 4320
Fan X., Carilli C. L., Keating B., 2006, Annu. Rev. Astron. Astrophys., 44, 415
Feroz F., Hobson M. P., Bridges M., 2009, MNRAS, 398, 1601
Finlator K. Özel F., Davé R., 2009, MNRAS, 393, 1090
Finlator K., Davé R., Özel F., 2011, ApJ, 743, 169
Finlator K., Muñoz J. A., Oppenheimer B. D., Oh S. P., Özel F., Davé R., 2013, MNRAS, 436, 1818
Finlator K., Thompson R., Huang S., Davé R., Zackrisson E., Oppenheimer B. D., 2015, MNRAS, 447, 2526
Foreman-Mackey D., 2016, J. Open Source Softw., 1, 2
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Gnedin N. Y., 2000, ApJ, 542, 535
Gnedin N. Y., 2014, ApJ, 793, 29
Greig B., Mesinger A., 2015, MNRAS, 449, 4246
Greig B., Mesinger A., 2017, MNRAS, 465, 4838
Hassan S., Davé R., Finlator K., Santos M. G., 2016, MNRAS, 457, 1550
Hinshaw G. et al., 2013, ApJS, 208, 19
Iliev I. T., Mellema G., Ahn K., Shapiro P. R., Mao Y., Pen U.-L., 2014, MNRAS, 439, 725
Iliev I. T., Santos M. G., Mesinger A., Majumdar S., Mellema G., 2015, PoS (AASKA14), Advancing Astrophysics with the Square Kilometre Array. SISSA, Trieste, p. 007
Jensen H. et al., 2013, MNRAS, 435, 460
Katz H., Kimm T., Sijacki D., Haehnelt M., 2016, MNRAS, preprint (arXiv:1612.01786)
Keating L. C., Haehnelt M. G., Becker G. D., Bolton J. S., 2014, MNRAS, 438, 1820
Kennicutt R. C. Jr., 1998, ARA&A, 36, 189
Kulkarni G., Choudhury T. R., Puchwein E., Haehnelt M. G., 2016, MNRAS, 463, 2583
McQuinn M., Lidz A., Zahn O., Dutta S., Hernquist L., Zaldarriaga M., 2007, MNRAS, 377, 1043
Majumdar S., Mellema G., Datta K. K., Jensen H., Choudhury T. R., Bharadwaj S., Friedrich M. M., 2014, MNRAS, 443, 2843
Mellema G., Iliev I. T., Pen U.-L., Shapiro P. R., 2006, MNRAS, 372, 679
Mesinger A., Furlanetto S., 2007, ApJ, 669, 663
Mesinger A., Furlanetto S., Cen R., 2011, MNRAS, 411, 955
Mitra S., Choudhury T. R., Ferrara A., 2011, MNRAS, 413, 1569
Mitra S., Choudhury T. R., Ferrara A., 2012, MNRAS, 419, 1480

Mitra S., Ferrara A., Choudhury T. R., 2013, MNRAS, 428, L1
Mitra S., Choudhury T. R., Ferrara A., 2015, MNRAS, 454, L76
Paranjape A., Choudhury T. R., Padmanabhan H., 2016, MNRAS, 460, 1801
Parsons A., Pober J., McQuinn M., Jacobs D., Aguirre J., 2012, ApJ, 753, 81
Pawlik A. H., Schaye J., 2008, MNRAS, 389, 651
Planck Collaboration XIII, 2016a, A&A, 594, A13
Planck Collaboration XLVII, 2016b, A&A, 596, A108
Pober J. C. et al., 2013, AJ, 145, 65
Pober J. C. et al., 2014, ApJ, 782, 66
Razoumov A. O., Norman M. L., Abel T., Scott D., 2002, ApJ, 572, 695
Santos M. G., Ferramacho L., Silva M. B., Amblard A., Cooray A., 2010, MNRAS, 406, 2421
Schaerer D., 2003, A&A, 397, 527
Sobacchi E., Mesinger A., 2014, MNRAS, 440, 1662
Thomas R. M. et al., 2009, MNRAS, 393, 32
van Haarlem M. P. et al., 2013, A&A, 556, A2
Zahn O., Lidz A., McQuinn M., Dutta S., Hernquist L., Zaldarriaga M., Furlanetto S. R., 2007, ApJ, 654, 12
Zahn O., Mesinger A., McQuinn M., Trac H., Cen R., Hernquist L. E., 2011, MNRAS, 414, 727
Zel'Dovich Y. B., 1970, A&A, 5, 84

APPENDIX A: REIONIZATION HISTORY CONVERGENCE TEST

As mentioned earlier, our new time-integrated EoR model is sensitive to the choice of simulation time-step dz, due to implementing time-integrated ionization condition (equation 5). The correct reionization history is however achieved for very small values of dz (ideally when dz goes to zero), which is computationally impossible. We here present a convergence test to support our choice of d$z = 0.125$. Fig. A1 shows the well-calibrated time-integrated model reionization history for different time-steps dz using a simulation box size of $L = 75$ Mpc and $N = 140^3$. It is clear that the reionization starts earlier (higher τ) when adopting larger steps (d$z = 1.0$ in black or 0.5 in yellow) than using smaller step values (d$z \leq 0.25$). This means using larger dz permits more ionizing photon production, and hence lower neutral fractions as seen in the beginning of the reionization.

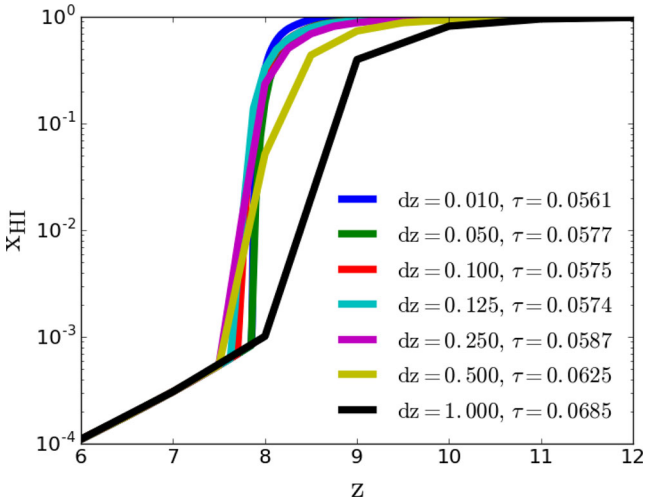

Figure A1. Reionization history convergence test in our new time-integrated EoR model using different time interval steps as quoted in the legend along with the corresponding optical depth. It is evident that our model is well converged for d$z \leq 0.125$.

Comparing the reionization history obtained with d$z = 0.125$ (cyan) versus that with d$z = 0.01$ (blue, the most correct result), we find that both models produce similar tau ($\Delta\tau \sim 0.001$) and the reionization history shape is identical. There is a very minor difference in the neutral fraction of $\Delta x_{\rm H\,I} \sim 0.01$ by end of reionization. Hence, we conclude that our time-integrated model is numerically well converged for d$z \leq 0.125$.

This paper has been typeset from a TEX/LATEX file prepared by the author.